Formale Sprachen und Komplexität, SS 18,
Prof. Dr. Volker Heun

Übungsblatt 4

Abgabe: bis Mo. 28.05.2018 8 Uhr

Institut für Informatik **Ludwig-Maximilians-Universität München**
Prof. Dr. Volker Heun

Formale Sprachen und Komplexität, SS 18
Übungsblatt 4

Abgabe: bis Mo. 28.05.2018 8 Uhr

Nach Bearbeitung dieses Übungsblattes sollten Sie:

	Check
Reguläre Ausdrücke zu einer informell beschriebenen Sprache erstellen können.	
Das Pumping-Lemma für reguläre Sprachen kennen.	
Mit Hilfe des Pumping-Lemmas für reguläre Sprachen kontrapositiv Beweise führen können.	
Mit der Relation R_L auf einer formalen Sprache L umgehen können.	
Den Satz von Myhill-Nerode kennen und anwenden können.	
Den Minimalautomaten aus einem gegebenen Automaten konstruieren können.	

Diese Ziele sind wichtige Hinweise für die Klausur!

Aufgabe 4-1 schriftlich bearbeiten
reguläre Ausdrücke

Für alle regulären Ausdrücke α_1, α_2, α_3 gilt $L\Big(\ ((\alpha_1|\alpha_2)\ |\ \alpha_3)\ \Big) = L\Big(\ (\alpha_1\ |\ (\alpha_2|\alpha_3))\ \Big)$. Deshalb können Sie in dieser Aufgabe die Schreiberleichterung $(\alpha_1|\alpha_2|\alpha_3)$ anstelle der voll geklammerten Formen benutzen, aber ansonsten **nur Konstrukte aus der Definition der regulären Ausdrücke.**

Sei L die Sprache der Literale für Gleitpunktkonstanten wie `+314.15926535e-2` in Programmiersprachen. Dabei soll gelten: Alle Vorzeichen sind optional. Führende Nullen sind erlaubt (brauchen also keine Sonderbehandlung).
Vor und hinter dem Dezimalpunkt muss jeweils mindestens eine Ziffer stehen, aber es dürfen beliebig viele sein. Der Exponent-Anteil ab `e` darf entfallen, aber wenn er vorhanden ist, muss er mindestens eine Ziffer enthalten.

Geben Sie einen regulären Ausdruck β an, so dass $L(\beta) = L$ ist.

Lösungsvorschlag:

$\Sigma = \{+, -, \textbf{.}, e, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$

Erste Idee, die aber **nicht** dem Formalismus entspricht:

$$\alpha = (0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9)$$
$$\beta = (\varepsilon|+|-)\, \alpha^+ \textbf{.}\, \alpha^+ \left(\varepsilon \;\middle|\; e\,(\varepsilon|+|-)\, \alpha^+\right)$$

Das so definierte β ist aus zwei Gründen kein regulärer Ausdruck wie im Buch definiert:

- Die Definition der regulären Ausdrücke enthält keinen $^+$-Operator.
  (Es wäre allerdings kein Problem, einen $^+$-Operator zu erlauben.)

  Man muss α^+ anders ausdrücken, zum Beispiel durch $\alpha\alpha^*$
- Die Definition der regulären Ausdrücke enthält keine Variablen wie α.
  (Würde man Variablen erlauben, könnte man damit nichtreguläre Sprachen definieren.)

  Man muss α durch Einsetzen beseitigen, was in diesem Fall möglich ist.

Endgültige Lösung

$$\beta = (\varepsilon|+|-)\,(0|1|2|3|4|5|6|7|8|9)\,(0|1|2|3|4|5|6|7|8|9)^* \textbf{.}\, (0|1|2|3|4|5|6|7|8|9)\,(0|1|2|3|4|5|6|7|8|9)^* \left(\varepsilon \;\middle|\; e\,(\varepsilon|+|-)\,(0|1|2|3|4|5|6|7|8|9)\,(0|1|2|3|4|5|6|7|8|9)^*\right)$$

Hinweise

Das Zeichen **.** hat keine Sonderbedeutung in diesem Formalismus.
(Es hätte eine Sonderbedeutung in „regulären Ausdrücken“ aus Programmiersprachen.)

Aufgabe 4-2 schriftlich bearbeiten
Pumping Lemma (regulär)

Es sei $\Sigma = \{a, b\}$ und für $w \in \Sigma^*$ sei $|w|_a$ die Anzahl der a's und $|w|_b$ die Anzahl der b's im Wort w.

Sei $L = \left\{ w \in \Sigma^* \;\middle|\; |w|_a < |w|_b \right\}$ die Menge aller Wörter, die mehr b's als a's enthalten.

Beweisen Sie mit Hilfe des Pumping-Lemmas, dass L nicht vom Typ 3 (regulär) ist.

Lösungsvorschlag:

Erst mal ein paar Beispiele, um L besser zu verstehen:

$abb,\ bab,\ bba,\ b,\ bbbb \quad \in L$

$aab,\ aba,\ baa,\ a,\ abab,\ \varepsilon \quad \notin L$

Behauptung: L ist nicht vom Typ 3 (regulär)

Beweis: Widerspruchsbeweis mit dem Pumping-Lemma

Annahme: L sei vom Typ 3 (regulär).
Sei n eine Pumping-Lemma-Zahl von L.
Betrachte $z = a^n b^{n+1} \in L$. Es gilt $|z| = 2n + 1 \geq n$.

Sei $z = uvw$ eine Zerlegung mit

$1 \leq |v| \qquad |uv| \leq n \qquad \forall i \in \mathbb{N}$ gilt $uv^iw \in L$

Für diese Zerlegung gilt $uvw = a^n b^{n+1}$ und $|uv| \leq n$.
Im Teilwort v kann also kein b vorkommen.
Also ist $v = a^k$ für ein $k \in \mathbb{N}$ mit $1 \leq k \leq n$.

Dann ist $uv^2w = a^{n+k}\, b^{n+1} \notin L$ Widerspruch.

Aufgabe 4-3 Äquivalenzrelation R_L und Pumping-Lemma

Sei $\Sigma = \{a, b, c\}$ und $L = \{b, c\}^* \cup \{a^k b^m c^m \mid k > 0,\ m > 0\}$

a) Zeigen Sie, dass der Index von R_L unendlich ist.

Definition für $x, y \in \Sigma^*$:
$x\ R_L\ y$ gdw. für alle $z \in \Sigma^*$ gilt $xz \in L \Leftrightarrow yz \in L$

Lösungsvorschlag:

Informelle Überlegung:

x, y	ε	c	c^2	c^3	c^4	$\dots$
abc	$\in L$	$\notin L$	$\notin L$	$\notin L$	$\notin L$	
ab^2c	$\notin L$	$\in L$	$\notin L$	$\notin L$	$\notin L$	
ab^3c	$\notin L$	$\notin L$	$\in L$	$\notin L$	$\notin L$	
ab^4c	$\notin L$	$\notin L$	$\notin L$	$\in L$	$\notin L$	
$\vdots$						$\ddots$

(Die Spalten $\varepsilon, c, c^2, c^3, c^4, \dots$ sind mit z überschrieben.)

Formaler Beweis:

Sei $\begin{cases} x = ab^ic \\ y = ab^jc \\ i \neq j \end{cases}$ sei $z = c^{i-1}$, dann ist $\begin{array}{ll} xz = ab^ic^i & \in L \\ yz = ab^jc^i & \notin L \end{array}$

Die unendlich vielen Wörter der Form ab^ic mit $i \in \mathbb{N} \setminus \{0\}$ stehen also paarweise nicht in der Relation R_L.

Das heißt, $Index(R_L) = \infty$.

Nach dem Satz von Myhill/Nerode (Buch S. 34) folgt daraus: L ist nicht vom Typ 3 (regulär).

b) Zeigen Sie, dass L die Pumping-Lemma-Bedingung erfüllt, das heißt:

$\exists n \in \mathbb{N}$ (Pumping-Lemma-Zahl von L) so dass
$\forall z \in L$ mit $|z| \geq n$ gilt
$\exists u,v,w \in \Sigma^*$ mit $z = uvw$ (Zerlegung von z) so dass

1. $1 \leq |v|$
2. $|uv| \leq n$
3. $\forall i \in \mathbb{N}$ gilt $uv^iw \in L$.

Lösungsvorschlag:

Wir zeigen, dass L die Pumping-Lemma-Bedingung für $n = 1$ erfüllt.

Beweis:
Für $z \in L$ mit $|z| \geq n = 1$ gibt es nur folgende Fälle:

Fall $z = \underbrace{}_{\varepsilon=u}\underbrace{b}_{v}\underbrace{\boxed{\in \{b,c\}^*}}_{w}$

Dann: $\forall i \in \mathbb{N}$ gilt $uv^iw = b^iw \in \{b,c\}^* \subseteq L$.

Fall $z = \underbrace{}_{\varepsilon=u}\underbrace{c}_{v}\underbrace{\boxed{\in \{b,c\}^*}}_{w}$

Dann: $\forall i \in \mathbb{N}$ gilt $uv^iw = c^iw \in \{b,c\}^* \subseteq L$.

Fall $z = \underbrace{}_{\varepsilon=u}\underbrace{a}_{v}\underbrace{\overbrace{\boxed{a\ \ldots\ a}}^{j}\overbrace{\boxed{b\ \ldots\ b}}^{m}\overbrace{\boxed{c\ \ldots\ c}}^{m}}_{w}$ $\quad j \geq 1,\ m \geq 1$

Dann: $\forall i \in \mathbb{N}$ gilt $uv^iw = a^iw = a^{i+j}\,b^mc^m \in L$

Fall $z = \underbrace{}_{\varepsilon=u}\underbrace{a}_{v}\underbrace{\overbrace{\boxed{b\ \ldots\ b}}^{m}\overbrace{\boxed{c\ \ldots\ c}}^{m}}_{w}$ $\quad m \geq 1$

Dann: für $i = 0$ gilt $uv^0w = w = b^mc^m \in \{b,c\}^* \subseteq L$.
$\forall i > 0$ gilt $uv^iw = a^iw = a^i\,b^mc^m \in L$.

Für jedes $z \in L$ mit $|z| \geq n = 1$ gibt es also eine Zerlegung $z = uvw$ mit 3.
Da in allen Fällen $|u| = 0$ und $|v| = 1$ ist, gilt für jede dieser Zerlegungen auch 1. und 2.

c) Was bedeutet dieses Ergebnis für das Pumping-Lemma?

Lösungsvorschlag:

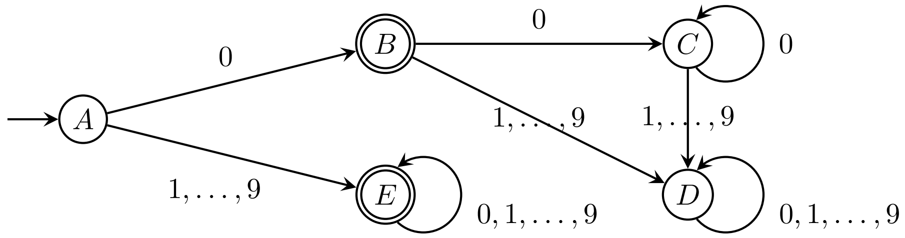
L ist nicht regulär, erfüllt aber trotzdem die Pumping-Lemma-Bedingung.

Also gilt die Gegenrichtung des Pumping-Lemmas nicht.
Es liefert nur eine notwendige Bedingung für die Regularität, aber keine hinreichende.

Pragmatischer:
Die Nichtregularität einer Sprache kann man oft durch Widerspruchsbeweis mit dem Pumping-Lemma zeigen, aber das klappt nicht immer. Es gibt nichtreguläre Sprachen, für die man mit dem Pumping-Lemma keinen Widerspruch konstruieren kann.

Aufgabe 4-4 schriftlich bearbeiten
Minimalautomat

Der folgende deterministische endliche Automat M akzeptiert die Menge aller natürlichen Zahlen in Dezimalschreibweise, aber ohne führende Nullen und ohne Vorzeichen. Sei $M = (Z, \Sigma, \delta, A, \{B, E\})$ mit $Z = \{A, B, C, D, E\}$ und $\Sigma = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ und δ gegeben durch den Zustandsgraph:



Konstruieren Sie den Minimalautomaten nach dem Verfahren aus dem Buch (S. 38). (Häufig wird dieses Verfahren als Table-Filling Algorithmus bezeichnet)

Lösungsvorschlag:

1. Tabelle (Matrix), Zeilen/Spalten indiziert mit Zuständen

Markierungen in der Tabelle bedeuten „verschiedenes Akzeptanzverhalten".

Diese Beziehung ist symmetrisch und irreflexiv.
Deshalb ist nur der Teil unter der Hauptdiagonalen relevant.

	A	B	C	D	E
A					
B					
C					
D					
E					

Zwei Zustände haben verschiedenes Akzeptanzverhalten, wenn es ein Wort $w \in \Sigma^*$ gibt, das ausgehend vom einen akzeptiert wird und ausgehend vom anderen nicht.

Dann können diese Zustände nicht zu einem verschmolzen werden.
Denn falls der Automat in diesen verschmolzenen Zustand käme und das Restwort w wäre, wäre nicht mehr unterscheidbar, ob er akzeptieren soll oder nicht.

2. Initiale Markierung

Alle Paare Endzustand/Nicht-Endzustand

(sie haben verschiedenes Akzeptanzverhalten für ε)

	A	B	C	D	E
A					
B	$\star$				
C		$\star$			
D		$\star$			
E	$\star$		$\star$	$\star$	

3. Vervollständigen der Tabelle

$\genfrac{}{}{0pt}{}{C}{A} \xrightarrow{0} \genfrac{}{}{0pt}{}{C}{B}$ und $\genfrac{}{}{0pt}{}{C}{B}$ bereits markiert: markiere $\genfrac{}{}{0pt}{}{C}{A}$

$\genfrac{}{}{0pt}{}{D}{A} \xrightarrow{0} \genfrac{}{}{0pt}{}{D}{B}$ und $\genfrac{}{}{0pt}{}{D}{B}$ bereits markiert: markiere $\genfrac{}{}{0pt}{}{D}{A}$

$\genfrac{}{}{0pt}{}{E}{B} \xrightarrow{0} \genfrac{}{}{0pt}{}{E}{C}$ und $\genfrac{}{}{0pt}{}{E}{C}$ bereits markiert: markiere $\genfrac{}{}{0pt}{}{E}{B}$

A					
B	⋆				
C	⋆	⋆			
D	⋆	⋆			
E	⋆	⋆	⋆	⋆	
	A	B	C	D	E

$$\left.\begin{array}{l} \genfrac{}{}{0pt}{}{D}{C} \xrightarrow{0} \genfrac{}{}{0pt}{}{D}{C} \text{ und } \genfrac{}{}{0pt}{}{D}{C} \text{ unmarkiert} \\ \genfrac{}{}{0pt}{}{D}{C} \xrightarrow{1...9} \genfrac{}{}{0pt}{}{D}{D} \text{ und } \genfrac{}{}{0pt}{}{D}{D} \text{ unmarkiert} \end{array}\right\} \genfrac{}{}{0pt}{}{D}{C} \text{ bleibt unmarkiert}$$

- Ein Paar kann markiert werden, sobald der Test für **ein** Element von Σ das Ergebnis „Zielpaar bereits markiert“ hat.
  Der Test für andere Elemente von Σ ist dann überflüssig.
- Ein Paar bleibt nur dann unmarkiert, wenn der Test für **alle** Elemente von Σ das Ergebnis „Zielpaar unmarkiert“ hat.

  Obendrein kann sich dieses Ergebnis ändern, wenn neue Marken hinzukommen. Der Test für unmarkierte Paare muss also so lange wiederholt werden, bis sich nichts mehr ändert.

Erläuterung zur Vervollständigung:

Wenn $\genfrac{}{}{0pt}{}{E}{C}$ markiert sind, haben sie verschiedenes Akzeptanzverhalten. Es gibt also ein Wort w (nämlich ε, aber man braucht es nicht zu kennen), das ausgehend vom einen Zustand akzeptiert wird, vom anderen nicht.

Wegen $\genfrac{}{}{0pt}{}{E}{B} \xrightarrow{0} \genfrac{}{}{0pt}{}{E}{C}$ haben auch $\genfrac{}{}{0pt}{}{E}{B}$ verschiedenes Akzeptanzverhalten, denn das Wort $0w$ wird ausgehend vom einen akzeptiert, vom anderen nicht.

Aus dem verschiedenen Akzeptanzverhalten von E und C folgt also verschiedenes Akzeptanzverhalten des Vorgängerpaars E und B.

4. Minimalautomat

Laut End-Markierung der Tabelle haben C und D gleiches Akzeptanzverhalten, alle anderen Paare verschiedenes.

Also werden C und D verschmolzen, der Rest bleibt wie in M.