TIN 2016/2017: Úloha 1

Dávid Mikuš
xmikus15@stud.fit.vutbr.cz

20. októbra 2016

1. Uvažujte jazyk $L_1 = \{a^i b^j c^k \mid (i \neq k) \vee (j = 2l, l \geq 0)\}$.

(a) Zostavte gramatiku G_1 takú, že $L(G_1) = L_1$.

(b) Akého typu (podľa Chomského hierachie jazykov) je G_1 a akého typu je L_1? Môžu sa tieto typy všeobecne líšiť? Svoje tvrdenie zdôvodnite (formálny dôkaz nie je požadovaný).

Riešenie

(a)

$G_1 = (N, \Sigma, P, S)$

$N = \{S, A, B, D, E, F, G\}$
$\Sigma = \{a, b, c\}$
$P:$
$S \to DCc \mid aE \mid AFC$
$A \to aA \mid \varepsilon$
$B \to bB \mid \varepsilon$
$C \to cC \mid \varepsilon$
$D \to aDc \mid B \mid \varepsilon$
$E \to aEG \mid B \mid \varepsilon$
$F \to bbF \mid \varepsilon$
$G \to c \mid \varepsilon$

(b) Gramatika G_1 je typu 2 pretože obsahuje pravidlá tvaru $A \to \alpha$ kde $A \in N$ a $\alpha \in (N \cup \Sigma)^*$. Typu 3 už nemôže byť pretože pravidlá porušuju tvar $A \to xB \mid x$ kde $A, B \in N$ a $x \in \Sigma^*$
Jazyk L_1 je typu 2, pretože už nemôže byť typom 3 (regulárny jazyk).
Tieto typy sa môžu všeobecne líšiť. Gramatika môže mať rovnaký typ ako jazyk alebo byť vyššieho typu v porovnaní s jazykom ktorý generuje. Napríklad bezkontextová gramatika (Typ 2) môže generovať regulárny jazyk (Typ 3).

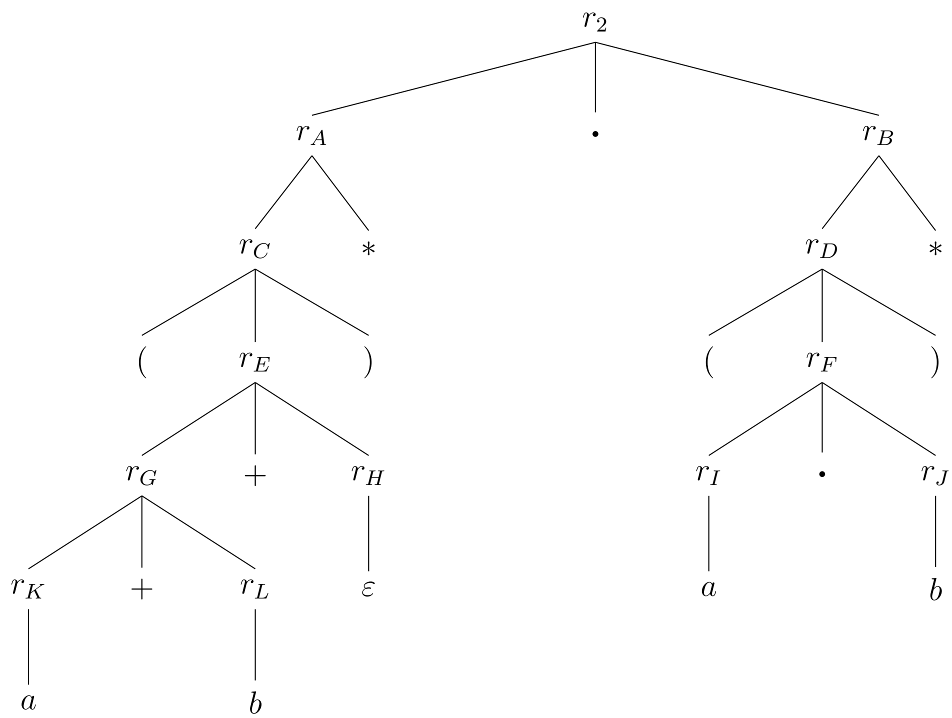
2. Uvažujte regulárny výraz $r_2 = (a + b + \varepsilon)^*(ab)^*$.

(a) Preveďte r_2 algoritmicky na redukovaný deterministický konečný automat M_2 (tj. RV $\rightarrow$ RKA $\rightarrow$ DKA $\rightarrow$ redukovaný DKA), prijímajúci jazyk popísaný výrazom r_2.

(b) Pre jazyk $L(M_2)$ určte počet tried ekvivalencie relácie $\sim_L$ (viz. Myhill-Nerodova veta) a vypíšte tieto triedy. Jednotlivé triedy môžete popísať napríklad konečným automatom, ktorý akceptuje všetky slová patriace do danej triedy.

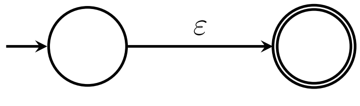
Riešenie

(a) Použijeme algoritmus 3.7 zo študijnej opory. Najskôr rozložíme r_2 na primitívne zložky a rozklad vyjadríme stromom.

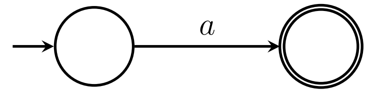


Obr. 1: Rozklad regulárneho výrazu r_2.

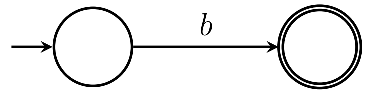
Následne zostrojíme ku každému regulárnemu výrazu príslušny konečny automat.



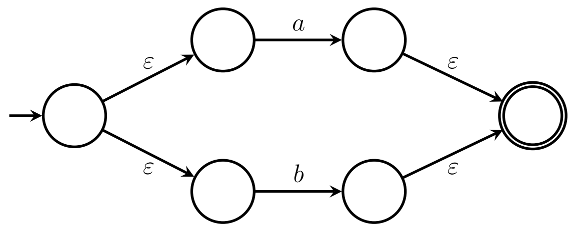
Obr. 2: RKA M_H ekvivalentný RV r_H.



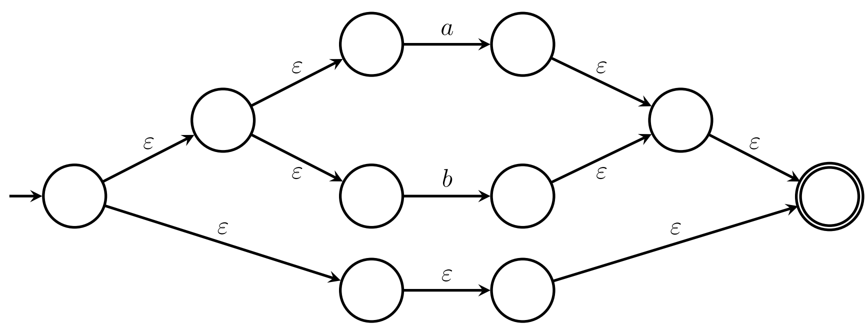
Obr. 3: RKA M_I, M_K ekvivalentný RV r_I, r_K.



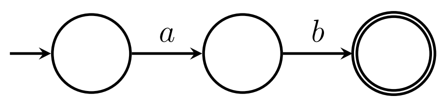
Obr. 4: RKA M_J, M_L ekvivalentný RV r_J, r_L.



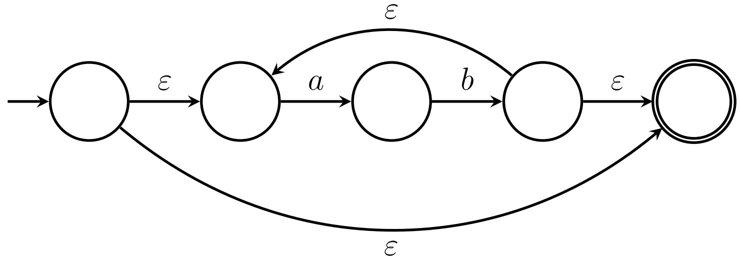
Obr. 5: RKA M_G ekvivalentný RV r_G.



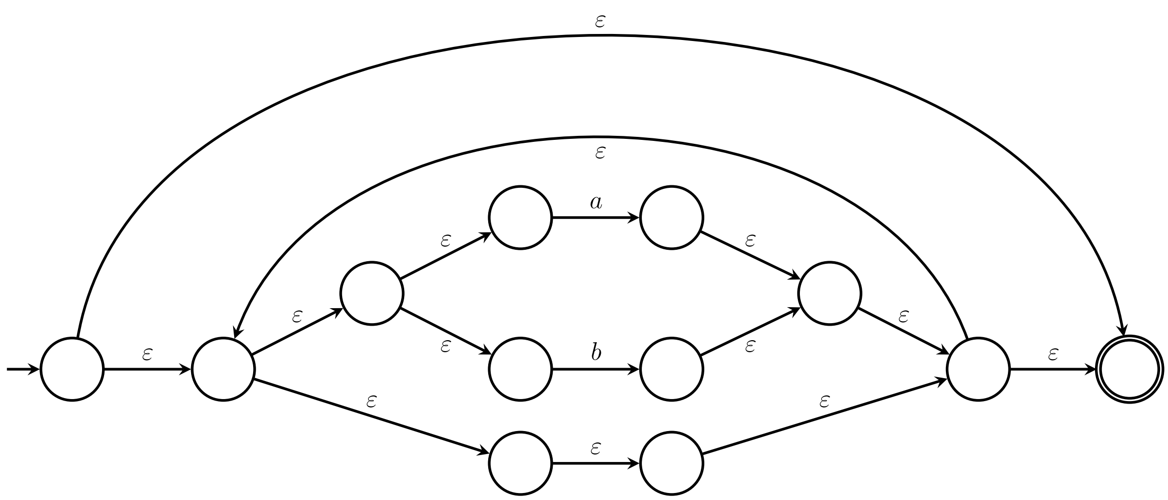
Obr. 6: RKA M_E, M_C ekvivalentný RV r_E, r_C.



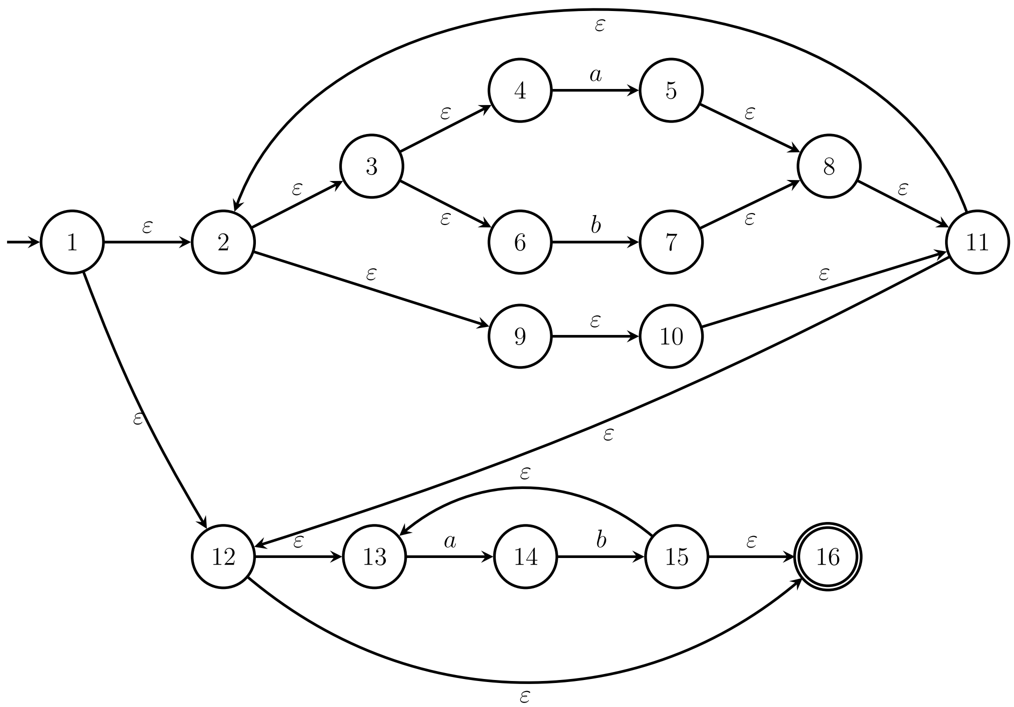
Obr. 7: RKA M_F, M_D ekvivalentný RV r_F, r_D.



Obr. 8: RKA M_B, ekvivalentný RV r_B.



Obr. 9: RKA M_A ekvivalentný RV r_A.



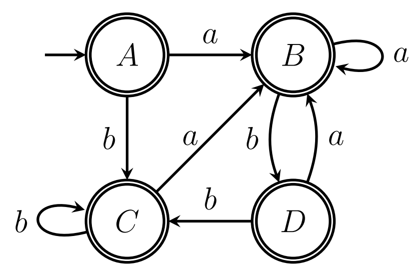
Obr. 10: RKA M_2 ekvivalentný RV r_2.

Podľa algoritmu 3.6 prevedieme RKA M_2 na DKA $M_D = (Q_D, \Sigma, \delta_D, q_D^0, F_D)$.

$q_D^0 = \varepsilon\text{-uzáver}(1) = \{1, 2, 3, 4, 6, 9, 10, 11, 12, 13, 16\} = A$

$\delta_D(A, a) = \varepsilon\text{-uzáver}(\{5, 14\}) = \{2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 16\} = B$
$\delta_D(A, b) = \varepsilon\text{-uzáver}(\{7\}) = \{2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 16\} = C$
$\delta_D(B, a) = \varepsilon\text{-uzáver}(\{5, 14\}) = B$
$\delta_D(B, b) = \varepsilon\text{-uzáver}(\{7, 15\}) = \{2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16\} = D$
$\delta_D(C, a) = \varepsilon\text{-uzáver}(\{5, 14\}) = B$
$\delta_D(C, b) = \varepsilon\text{-uzáver}(\{7\}) = C$
$\delta_D(D, a) = \varepsilon\text{-uzáver}(\{5, 14\}) = B$
$\delta_D(D, b) = \varepsilon\text{-uzáver}(\{7\}) = C$

$Q_D = \{A, B, C, D\}$
$F_D = \{A, B, C, D\}$

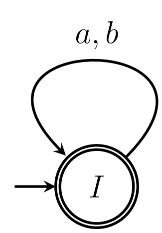


Obr. 11: DKA M_D.

Pomocou algoritmu 3.5 vykonáme prevod z DKA na redukovaný DKA.

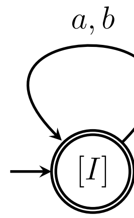
$\overset{0}{\equiv}$	δ_U	a	b
$I:$	A	B_I	C_I
	B	B_I	D_I
	C	B_I	C_I
	D	B_I	C_I

$\overset{0}{\equiv} = \{\{A, B, C, D\}\}$



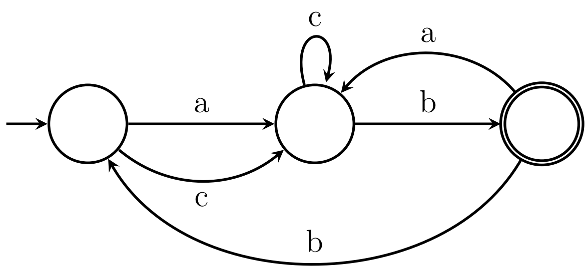
Obr. 12: Redukovaný DKA

(b) Počet tried ekvivalencie relácie $\sim_L$ je len jedná jediná a to $L^{-1}([I]) = (a+b)^*$.



Obr. 13: KA príjmajúci jazyk $L^{-1}([I]) = (a+b)^*$.

3. Uvažujte NKA M_3 nad abecedou $\Sigma = \{a, b, c\}$ z obrázku 14:



Obr. 14: NKA M_3

K tomuto automatu zostrojte sústavu rovníc nad regulárnymi výrazmi a jej riešením zostavte ekvivalentný regulárny výraz.

Riešenie

Označíme si jednotlive stavy KA z ľava doprava písmenami X, Y a Z a zostrojíme sústavu rovníc.

$$X = aY + cY \quad (1)$$
$$Y = cY + bZ \quad (2)$$
$$Z = bX + aY + \varepsilon \quad (3)$$

Z výrazu X vytkneme Y a na výraz Y použijeme vetu 3.14, ktorá hovorí že najmenším pevným bodom rovnice $X = pX + q$ je $X = p^*q$.

$$X = (a + c)Y \quad (4)$$
$$Y = c^*bZ \quad (5)$$

Výraz Y dosadíme do výrazov X a Z.

$$X = (a + c)c^*bZ \quad (6)$$
$$Z = bX + ac^*bZ + \varepsilon \quad (7)$$
$$= (ac^*b)^*(bX + \varepsilon) \quad (8)$$

Následne výraz Z dosadíme do X a upravujeme.

$$X = (a + c)c^*b(ac^*b)^*(bX + \varepsilon) \quad (9)$$
$$= (a + c)c^*b(ac^*b)^*bX + (a + c)c^*b(ac^*b)^* \quad (10)$$
$$= ((a + c)c^*b(ac^*b)^*b)^*(a + c)c^*b(ac^*b)^* \quad (11)$$

Výraz X je ekvivalentný regulárny výraz k automatu M_3

4. Majme jazyk $L_1 \subseteq \{a, b, c\}^*$ a jazyk $L_2 \subseteq \{0, 1\}^*$ a funkciu $bin : \{a, b, c\}^* \to \{0, 1\}$ definovanú následovne:

- $bin(\varepsilon) = \varepsilon, bin(a) = 00, bin(b) = 01$ a $bin(c) = 11$.
- Pre $|w| > 1$ platí: $(w = xw' \wedge |x| = 1) \Rightarrow bin(w) = bin(x)bin(w')$

Navrhnite a formálne popíšte algoritmus, ktorý má na vstupe dva konečné automaty $M_1 = (Q_1, \{a, b, c\}, \delta_1, q_1^0, F_1)$ a $M_2 = (Q_2, \{0, 1\}, \delta_2, q_2^0, F_1)$ (môžu byť nedeterministické), a jeho výstupom bude konečný automat M_f taký, že $L(M_f) = \{w | w \in L(M_1) \wedge bin(w) \in L(M_2)\}$.

Riešenie

$M_f = (Q_f, \Sigma_f, \delta_f, q_f^0, F_f)$

$Q_f = Q_1 \times Q_2$
$\Sigma_f = \{a, b, c\}$
$\delta_f : \forall q_1, q_2 \in Q_1 \forall q_3, q_4, q_5 \in Q_2 \forall x \in \Sigma_f :$

$$\begin{gathered}(q_2, q_5) \in \delta_f((q_1, q_3), x) \\ \Longleftrightarrow \\ (q_2 \in \delta_1(q_1, a) \wedge q_4 \in \delta_2(q_3, 0) \wedge q_5 \in \delta_2(q_4, 0)) \\ \vee \\ (q_2 \in \delta_1(q_1, b) \wedge q_4 \in \delta_2(q_3, 0) \wedge q_5 \in \delta_2(q_4, 1)) \\ \vee \\ (q_2 \in \delta_1(q_1, c) \wedge q_4 \in \delta_2(q_3, 1) \wedge q_5 \in \delta_2(q_4, 1))\end{gathered}$$

$q_f^0 = (q_1^0, q_2^0)$
$F_f = F_1 \times F_2$

5. Majme jazyk $L = \{a^i b^j c^k | i > k \wedge j \geq 1\}$. Je jazyk L regulárny? Dokážte alebo vyvráťte.

Riešenie

Predpokladajme, že jazyk L je regulárny jazyk. Potom platí, že

$$
\begin{aligned}
\exists k > 0 : \forall w \in L : |w| \geq k \Rightarrow \exists x, y, z \in \Sigma^* : \; & w = xyz \wedge \\
& y \neq \varepsilon \wedge \\
& |xy| \leq k \wedge \\
& \forall i \geq 0 : xy^i z \in L
\end{aligned}
$$

Zvolíme si reťazec w:

$$
\begin{aligned}
w &= a^k b c^{k-1}, w \in L \\
|w| &= 2k \geq k
\end{aligned}
$$

Reťazec $xy^i z \in L$

$$
\begin{aligned}
\exists x, y, z \in \Sigma^* : |xy| \leq k \\
x = a^{l_1}, y = a^{l_2}, \#_a(z) = l_3 \\
l_1, l_2 > 0, l_3 \geq 0 \\
\#_a(w) = l_1 + l_2 + l_3 = k
\end{aligned}
$$

Kedže $l_3 \geq 0$ tak potom $l_1 + l_2 \leq k$
Zvolíme si $i = 0$

$$
\begin{aligned}
xy^0 z = xz = a^{l_1} a^{l_3} b c^{k-1} \\
|a^{l_1} a^{l_3}| = k - l_2, |c^{k-1}| = k - 1
\end{aligned}
$$

Zo zadania by malo platiť že $\#_a(w) > \#_c(w)$

$$
\begin{aligned}
\#_a(xy^0 z) &= l_1 + l_3 \\
\#_c(xy^0 z) &= k - 1 \\
l_1 + l_3 &> k - 1 \\
k - l_2 &> k - 1 \qquad l_2 > 0 \\
l_2 &< 1
\end{aligned}
$$

Kedže nám vyšlo $l_2 < 1 \wedge l_2 > 0$ pričom $l_2 \in N$, tak takéto l_2 neexistuje, takže $xy^0 z \notin L$, došli sme k sporu. Tým pádom naš predpoklad neplatí a jazyk L nie je regulárny jazyk.

Dávid Mikuš
xmikus15

6. Dokážte, že pre každý regulárny jazyk existuje jednoznačná gramatika (definícia jednoznačnej gramatiky – viz slidy 4, strana 11).

Riešenie

(a) Každý regulárny jazyk L sa dá špecifikovať pomocou konečneho automatu M tak že $L(M) = L$. Veta 3.6

(b) Konečný automat M sa dá previesť na regulárnu gramatiku G, tak že $L(G) = L(M)$. Veta 3.7

(c)
   i. Predpokladajme že gramatika G je viacznačná. $G = (Q_G, \Sigma_G, P, q_G^0)$
   ii. Podľa definície 4.5 je gramatika G viacznačná ak generuje aspoň jednu viacznačnú vetu w. Veta w je generovaná gramatikou G ku ktorej existujú aspoň 2 rôzne derivačné stromy s koncovými uzlami tvoriacu vetu w.
   iii. Z dôkazu pre vetu 3.7 vieme že KA M sa dá previesť na DKA N. $N = (Q_N, \Sigma_N, \delta_N, q_N^0, F_N)$
   iv. Z (c)ii teda platí $\forall r \in \Sigma_G \exists q \in Q_G \exists a \in \Sigma_G : |q \to ar| > 1$, potom platí $\exists q \in Q_N \exists a \in \Sigma_N : |\delta(q, a)| > 1$, čiže k odvodzovaciemu pravidlu gramatiky existuje ekvivalentný prechod v konečnom automate.
   v. Kedže na základe (c)iii musí pre KA existovať viac ako jeden prechod pre určitý stav a symbol, tak dochádzame k sporu. Kedže na základe definície DKA, môže existovať maximálne jeden takýto prechod. $\forall q \in Q_N \forall a \in \Sigma_N : |\delta(q, a)| \leq 1$
   vi. Gramatika tedy neni viacznačná. Definícia 4.5 hovorí že ak gramatika neni viacznačná tak je potom jednoznačná.

(d) Gramatika G je jednoznačná. Takže pre každý regulárny jazyk existuje jednoznačná gramatika.