1. Uvažte jazyk $L_1 = \{a^i b^j c^k \mid (i = 2j \vee j = 3k) \wedge i, j, k \geq 0\}$.

   (a) Sestavte gramatiku G_1 takovou, že $L(G_1) = L_1$.

   $G_1 = (\{S, X, Y, U, V\}, \{a, b, c\}, P, S)$ s pravidly:

$$
\begin{array}{lcl}
S & \rightarrow & X \mid Y \mid XU \mid VY \\
X & \rightarrow & aaXb \mid \epsilon \\
Y & \rightarrow & bbbYc \mid \epsilon \\
U & \rightarrow & Uc \mid \epsilon \\
V & \rightarrow & aV \mid \epsilon
\end{array}
$$

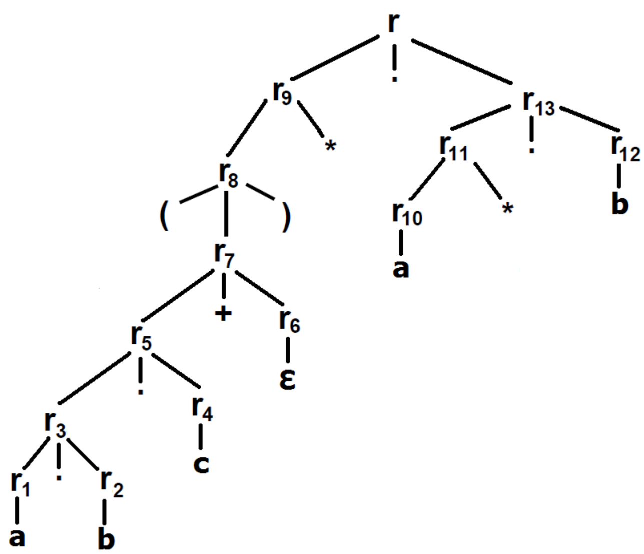
   (b) Jakého typu (dle Chomského hierarchie jazyků) je G_1 a jakého typu je L_1? Mohou se tyto typy obecně lišit? Svoje tvrzení zdůvodněte (formální důkaz není požadován).

   - G_1 je typu 2 (bezkontextová gramatika), protože se na levých stranách jejích přepisovacích pravidel vyskytuje vždy jen jeden nonterminální symbol a nic dalšího a zároveň se na některých pravých stranách vyskytují pouze nonterminály.
   - L_1 je typu 2, protože gramatika G_1, která ho generuje, je typu 2, a zároveň nemůže být generován gramatikou typu 3.
   - Platí, že $\mathcal{L}_3 \subseteq \mathcal{L}_2 \subseteq \mathcal{L}_1 \subseteq \mathcal{L}_0$, kde $\mathcal{L}_i$ značí třídu všech jazyků typu i. Tudíž gramatika G_i typu i může generovat jazyk třídy $\mathcal{L}_i$ nebo jazyk některé konkrétnější třídy $\mathcal{L}_{j>i}$. Gramatika a jazyk, který generuje, mohou tedy být obecně jiného typu.

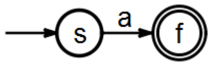
2. Uvažte regulární výraz $r = (abc + \epsilon)^* a^* b$.

   (a) Převeďte r algoritmicky na redukovaný deterministický konečný automat M (tj. RV $\rightarrow$ RKA $\rightarrow$ DKA $\rightarrow$ redukovaný DKA), přijímající jazyk popsaný výrazem r.

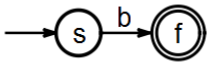
   - Převod regulárního výrazu r na rozšířený konečný automat.
   - Rozklad regulárního výrazu vyjádříme stromem:



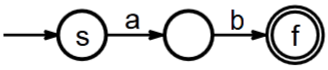
- Regulárnímu výrazu $r_1 = a$ přísluší konečný automat M_1:



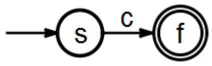
- Regulárnímu výrazu $r_2 = b$ přísluší konečný automat M_2:



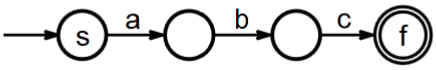
- Regulárnímu výrazu $r_3 = ab$ přísluší konečný automat M_3:



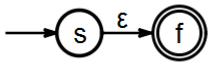
- Regulárnímu výrazu $r_4 = c$ přísluší konečný automat M_4:



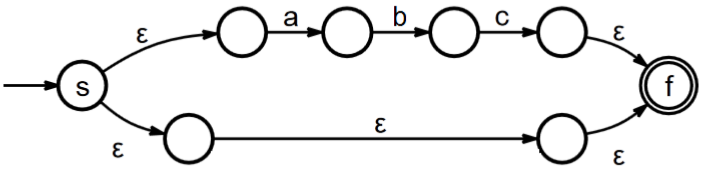
- Regulárnímu výrazu $r_5 = abc$ přísluší konečný automat M_5:



- Regulárnímu výrazu $r_6 = \epsilon$ přísluší konečný automat M_6:

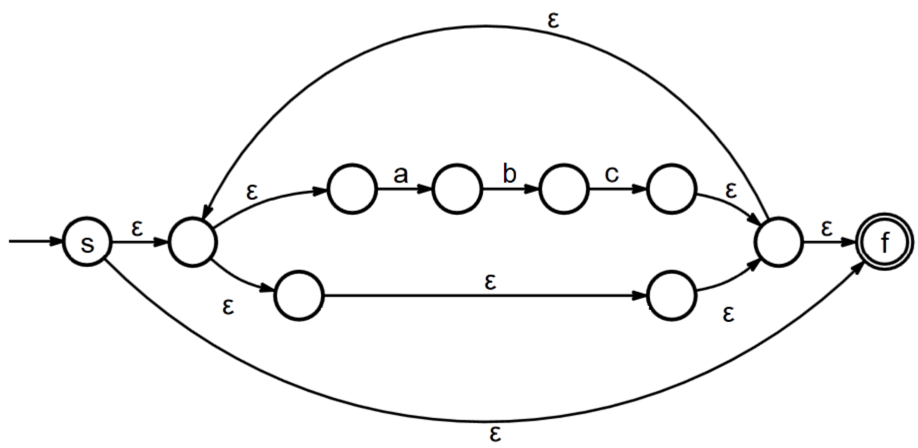


- Regulárnímu výrazu $r_7 = abc + \epsilon$ přísluší konečný automat M_7.

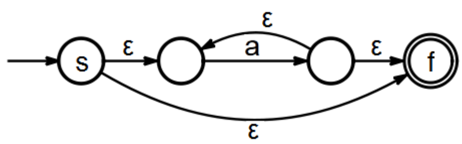


- Regulárnímu výrazu $r_8 = (abc + \epsilon)$ přísluší konečný automat M_7.

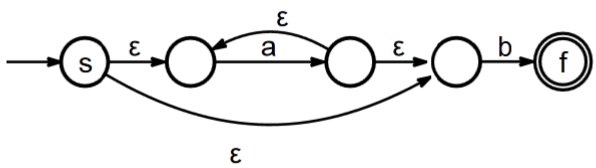
- Regulárnímu výrazu $r_9 = (abc + \epsilon)^*$ přísluší konečný automat M_9:



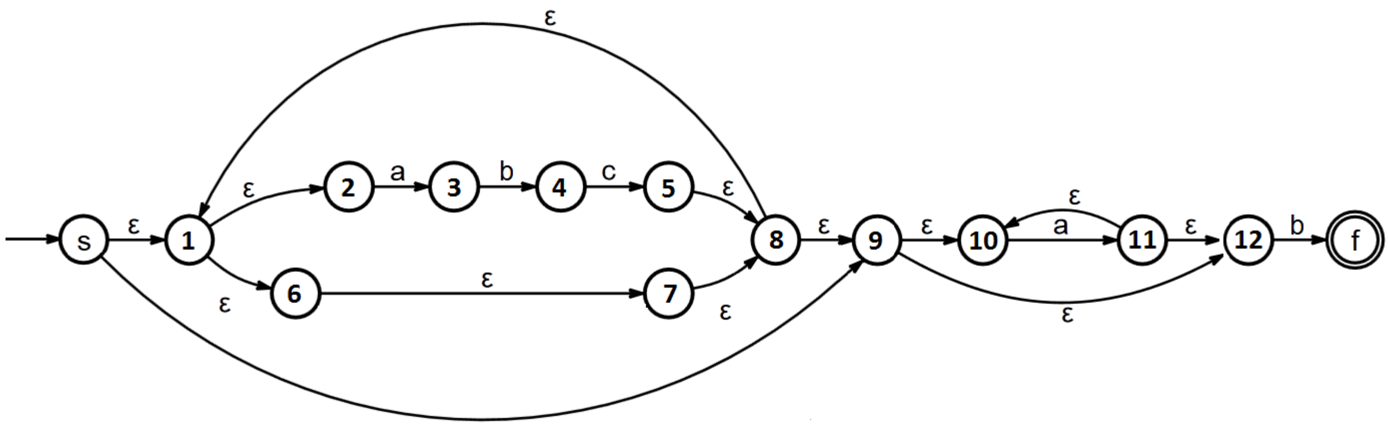
- Regulárnímu výrazu $r_{10} = r_1 = a$ přísluší konečný automat M_1.
- Regulárnímu výrazu $r_{11} = a^*$ přísluší konečný automat M_{11}:



- Regulárnímu výrazu $r_{12} = r_2 = b$ přísluší konečný automat M_2.
- Regulárnímu výrazu $r_{13} = a^*b$ přísluší konečný automat M_{13}:

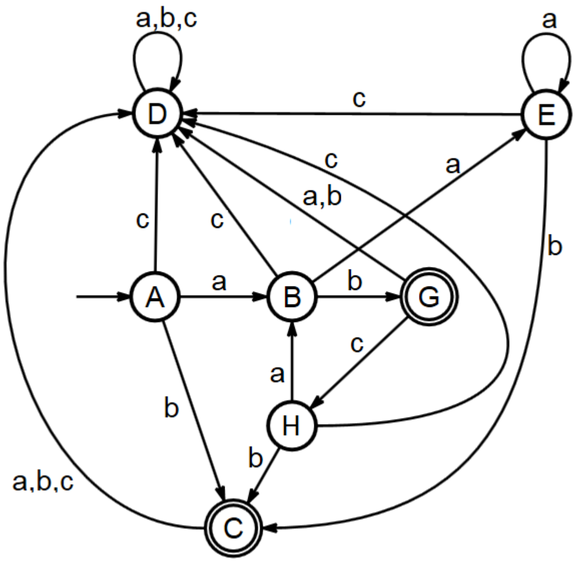


- Výslednému regulárnímu výrazu $r = (abc + \epsilon)^*a^*b$ přísluší rozšířený konečný automat M_R:



- Převedeme rozšířený konečný automat M_R na úplný deterministický konečný automat $M_D = (Q, \{a, b, c\}, \delta, A, F)$.

- Počáteční stav M_D = ϵ-uzávěr$(s) = \{s, 1, 2, 6, 7, 8, 9, 10, 12\} = A$.
- $\delta(A, a) = \epsilon$-uzávěr$(\{3, 11\}) = \{3, 10, 11, 12\} = B$.
- $\delta(A, b) = \epsilon$-uzávěr$(\{f\}) = \{f\} = C$.
- $\delta(A, c) = \epsilon$-uzávěr$(\emptyset) = \emptyset = D$.
- $\delta(B, a) = \epsilon$-uzávěr$(\{11\}) = \{10, 11, 12\} = E$.
- $\delta(B, b) = \epsilon$-uzávěr$(\{4, f\}) = \{4, f\} = G$.
- $\delta(B, c) = \epsilon$-uzávěr$(\emptyset) = \emptyset = D$.
- $\delta(C, a) = \epsilon$-uzávěr$(\emptyset) = \emptyset = D$.
- $\delta(C, b) = \epsilon$-uzávěr$(\emptyset) = \emptyset = D$.
- $\delta(C, c) = \epsilon$-uzávěr$(\emptyset) = \emptyset = D$.
- $\delta(D, a) = \epsilon$-uzávěr$(\emptyset) = \emptyset = D$.
- $\delta(D, b) = \epsilon$-uzávěr$(\emptyset) = \emptyset = D$.
- $\delta(D, c) = \epsilon$-uzávěr$(\emptyset) = \emptyset = D$.
- $\delta(E, a) = \epsilon$-uzávěr$(\{11\}) = \{10, 11, 12\} = E$.
- $\delta(E, b) = \epsilon$-uzávěr$(\{f\}) = \{f\} = C$.
- $\delta(E, c) = \epsilon$-uzávěr$(\emptyset) = \emptyset = D$.
- $\delta(G, a) = \epsilon$-uzávěr$(\emptyset) = \emptyset = D$.
- $\delta(G, b) = \epsilon$-uzávěr$(\emptyset) = \emptyset = D$.
- $\delta(G, c) = \epsilon$-uzávěr$(\{5\}) = \{1, 2, 5, 6, 7, 8, 9, 10, 12\} = H$.
- $\delta(H, a) = \epsilon$-uzávěr$(\{3, 11\}) = \{3, 10, 11, 12\} = B$.
- $\delta(H, b) = \epsilon$-uzávěr$(\{f\}) = \{f\} = C$.
- $\delta(H, c) = \epsilon$-uzávěr$(\emptyset) = \emptyset = D$.
- Výsledná množina stavů $Q = \{A, B, C, D, E, G, H\}$
- Množina koncových stavů $F = \{C, G\}$
- Výsledný úplný deterministický konečný automat M_D:



- Převod úplného deterministického konečného automatu M_D na výsledný redukovaný deterministický konečný automat M.
- M_D neobsahuje nedostupné stavy. Odstraníme nerozlišitelné stavy:

$\overset{0}{\equiv}$	δ	a	b	c
$I:$	C	D_{II}	D_{II}	D_{II}
	G	D_{II}	D_{II}	H_{II}
$II:$	A	B_{II}	C_I	D_{II}
	B	E_{II}	G_I	D_{II}
	D	D_{II}	D_{II}	D_{II}
	E	E_{II}	C_I	D_{II}
	H	B_{II}	C_I	D_{II}

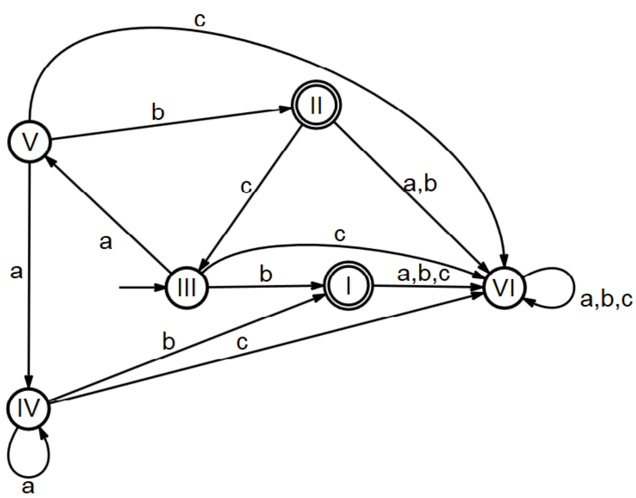
$\overset{1}{\equiv}$	δ	a	b	c
$I:$	C	D_{III}	D_{III}	D_{III}
	G	D_{III}	D_{III}	H_{II}
$II:$	A	B_{II}	C_I	D_{III}
	B	E_{II}	G_I	D_{III}
	E	E_{II}	C_I	D_{III}
	H	B_{II}	C_I	D_{III}
$III:$	D	D_{III}	D_{III}	D_{III}

$\overset{2}{\equiv}$	δ	a	b	c
$I:$	C	D_{IV}	D_{IV}	D_{IV}
$II:$	G	D_{IV}	D_{IV}	H_{III}
$III:$	A	B_{III}	C_I	D_{IV}
	B	E_{III}	G_{II}	D_{IV}
	E	E_{III}	C_I	D_{IV}
	H	B_{III}	C_I	D_{IV}
$IV:$	D	D_{IV}	D_{IV}	D_{IV}

$\overset{3}{\equiv}$	δ	a	b	c
$I:$	C	D_V	D_V	D_V
$II:$	G	D_V	D_V	H_{III}
$III:$	A	B_{IV}	C_I	D_V
	E	E_{III}	C_I	D_V
	H	B_{IV}	C_I	D_V
$IV:$	B	E_{III}	G_{II}	D_V
$V:$	D	D_V	D_V	D_V

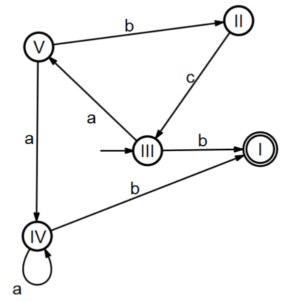
$\overset{3}{\equiv}$	δ	a	b	c
$I:$	C	D_{VI}	D_{VI}	D_{VI}
$II:$	G	D_{VI}	D_{VI}	H_{III}
$III:$	A	B_V	C_I	D_{VI}
	H	B_V	C_I	D_{VI}
$IV:$	E	E_{IV}	C_I	D_{VI}
$V:$	B	E_{IV}	G_{II}	D_{VI}
$VI:$	D	D_{VI}	D_{VI}	D_{VI}

- Výsledný redukovaný deterministický konečný automat M:

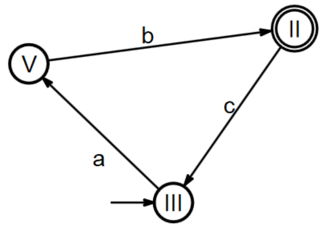


(b) Pro jazyk $L(M)$ určete počet tříd ekvivalence relace $\sim_L$ (viz. Myhill-Nerodova věta) a vypište tyto třídy. Jednotlivé třídy můžete popsat například konečným automatem, který akceptuje všechna slova patřící do dané třídy.

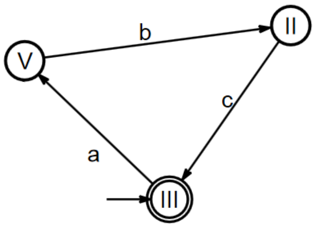
Celkem 6 tříd:



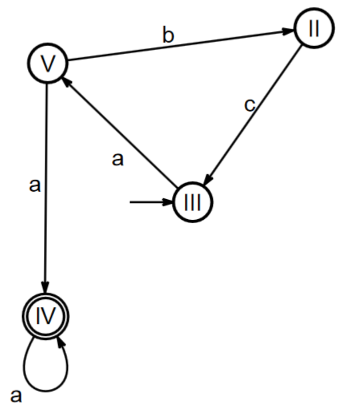
Obrázek 1: $L^{-1}(I)$



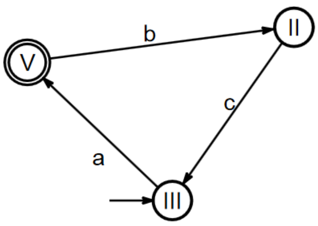
Obrázek 2: $L^{-1}(II)$



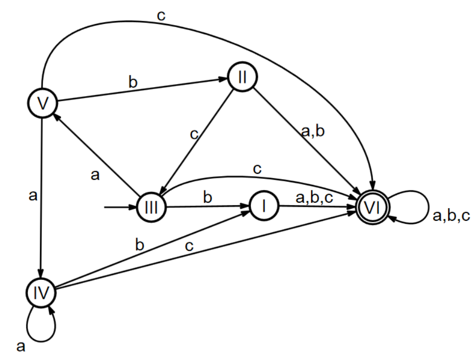
Obrázek 3: $L^{-1}(III)$



Obrázek 4: $L^{-1}(IV)$

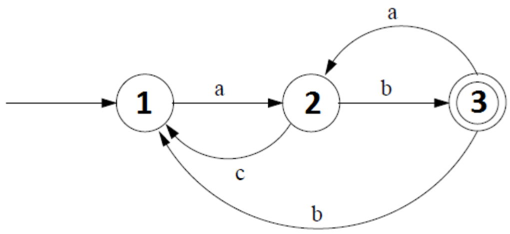


Obrázek 5: $L^{-1}(V)$



Obrázek 6: $L^{-1}(VI)$

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



Obrázek 7: NKA M_3

Řešením rovnic nad regulárními výrazy sestavte k tomuto automatu ekvivalentní regulární výraz.

Sestavíme soustavu rovnic nad regulárními výrazy pro M_3:

$X_1 = aX_2$
$X_2 = bX_3 + cX_1$
$X_3 = \epsilon + aX_2 + bX_1$

Regulárnímu výrazu ekvivalentnímu automatu M_3 odpovídá řešení této soustavy pro X_1.

Dosadíme do třetí rovnice za aX_2 podle první rovnice:

$X_3 = \epsilon + X_1 + bX_1$

Dosadíme výsledek do druhé rovnice za X_3:

$X_2 = b(\epsilon + X_1 + bX_1) + cX_1 = b + bX_1 + bbX_1 + cX_1$

Dosadíme výsledek do první rovnice za X_2:

$X_1 = a(b + bX_1 + bbX_1 + cX_1) = ab + abX_1 + abbX_1 + acX_1 = (ab + abb + ac)X_1 + ab$

Nejmenším pevným bodem rovnice nad regulárními výrazy ve tvaru $X = pX + q$ je $X = p^*q$, řešení pro X_1 a hledaný regulární výraz tedy je:

$X_1 = (ab + abb + ac)^*ab$

4. Mějme funkci $\Phi : \Sigma^* \to (\Sigma \cup \{\bullet\})^*$, kde $\bullet \notin \Sigma$, která každý třetí symbol ve slově nahradí symbolem $\bullet$. Formálně je Φ definována následujícím předpisem:

   - $\Phi(\epsilon) = \epsilon$
   - $\Phi(a_1) = a_1$, kde $a_1 \in \Sigma$
   - $\Phi(a_1 a_2) = a_1 a_2$, kde $a_1, a_2 \in \Sigma$
   - $\Phi(a_1 a_2 a_3 u) = a_1 a_2 \bullet \Phi(u)$, kde $a_1, a_2, a_3 \in \Sigma, u \in \Sigma^*$

   Například: $\Phi(aaa) = aa\bullet, \Phi(abcde) = ab \bullet de, \Phi(abcdef) = ab \bullet de\bullet$

   Navrhněte a formálně popište algoritmus, který má na vstupu konečný automat $M_1 = (Q, \Sigma, \delta, q_0, F)$ (může být nedeterministický), a jehož výstupem bude konečný automat M_2 takový, že $L(M_2) = \{\Phi(w) \mid w \in L(M_1)\}$.

   Algoritmus:

   Vstup: Konečný automat $M_1 = (Q_1, \Sigma_1, \delta_1, q_{01}, F_1)$ (může být i nedeterministický)

   Výstup: Konečný automat $M_2 = (Q_2, \Sigma_2, \delta_2, q_{02}, F_2)$ takový, že $L(M_2) = \{\Phi(w) \mid w \in L(M_1)\}$

   Metoda:

   - Polož $Q = \{0, 1, 2\}$
   - Polož $Q_2 = Q_1 \times Q$
   - Polož $\Sigma_2 = \Sigma_1 \cup \{\bullet\}$
   - Zkonstruuj $\delta_2 : Q_2 \times \Sigma_2 \to 2^{Q_2}$ tak, že:
     $\forall q_1, q_2 \in Q_1 \; \forall q_3, q_4 \in Q \; \forall a \in \Sigma_2 :$
     $(q_2, q_4) \in \delta_2((q_1, q_3), a) \Leftrightarrow$
     $(a \neq \bullet \wedge q_2 \in \delta_1(q_1, a) \wedge ((q_3 = 0 \wedge q_4 = 1) \vee (q_3 = 1 \wedge q_4 = 2))) \vee$
     $(a = \bullet \wedge \exists b \in \Sigma_1 : q_2 \in \delta_1(q_1, b) \wedge q_3 = 2 \wedge q_4 = 0)$
   - Polož $q_{02} = (q_{01}, 0)$
   - Polož $F_2 = F_1 \times Q$

5. Mějme jazyk $L = \{w \mid w \in \{a, b, c, d\}^* \wedge \#_a(w) = \#_b(w) \wedge \#_c(w) = \#_d(w)\}$, kde $\#_x(w)$ je počet symbolů x ve slově w. Je jazyk L regulární? Dokažte nebo vyvraťte.

   Jazyk L je nekonečný. Předpokládejme, že jazyk L je regulární. Pak podle Pumping lemma pro regulární jazyky platí:
   $\exists k > 0 : \forall w \in L : |w| \geq k \Rightarrow \exists x, y, z \in \Sigma^* : w = xyz \wedge |y| > 0 \wedge |xy| \leq k \wedge \forall i \geq 0 : xy^iz \in L$

   Uvažme libovolné $k > 0$ takové, že:
   $\forall w \in L : |w| \geq k \Rightarrow \exists x, y, z \in \Sigma^* : w = xyz \wedge |y| > 0 \wedge |xy| \leq k \wedge \forall i \geq 0 : xy^iz \in L$

   Zvolme $w = a^kb^k \in L$, $|w| = 2k > k$ a tedy z výše uvedeného:
   $\exists x, y, z \in \Sigma^* : w = a^kb^k = xyz \wedge |y| > 0 \wedge |xy| \leq k \wedge \forall i \geq 0 : xy^iz \in L$

   Uvažme libovolné $x, y, z \in \Sigma^*$ takové, že:
   $w = a^kb^k = xyz \wedge |y| > 0 \wedge |xy| \leq k \wedge \forall i \geq 0 : xy^iz \in L$

   Zvolme $i = 2k > 0$, tedy:
   $w = a^kb^k = xyz \wedge |y| > 0 \wedge |xy| \leq k \wedge v = xy^{2k}z \in L$

   Z toho je zřejmé, že řetězec xy může obsahovat pouze symboly a, proto pro řetězec v musí platit, že:
   $\#_a(v) \geq 2k \wedge \#_b(v) = k$,

   tudíž
   $\#_a(v) > \#_b(v)$,

   takže
   $v = xy^{2k}z \notin L$,

   což je SPOR. Jazyk L tudíž není regulární.

6. Dokažte, že pro každý regulární jazyk existuje jednoznačná gramatika (definice jednoznačné gramatiky - viz. slidy 4, strana 11).

   Podle definice **4.5** ze slidů 4 je jednoznačná gramatika taková gramatika, ve které existuje pro každou větu, kterou v ní lze vygenerovat, jen jeden derivační strom s koncovými uzly tvořícími tuto větu.

   Dále podle studijní opory platí:

   - Každý regulární jazyk lze vyjádřit ekvivalentním regulárním výrazem.
   - Každý regulární výraz lze převést na ekvivalentní deterministický konečný automat (např. algoritmy **3.7** a **3.6** ve studijní opoře).
   - Podle důkazu věty **3.7** ve studijní opoře lze každý deterministický konečný automat $M = (Q, \Sigma, \delta, q_0, F)$ vyjádřit ekvivalentní gramatikou $G = (Q, \Sigma, P, q_0)$ typu 3 s pravidly definovanými takto:

     (a) Je-li $\delta(q, a) = r$, pak P obsahuje pravidlo $q \to ar$
     (b) Je-li $p \in F$, pak P obsahuje pravidlo $p \to \epsilon$

     kde $q, r, p \in Q$, $a \in \Sigma$

   - Z bodu (a) plyne, že G nemůže obsahovat žádná dvě pravidla tvaru:

     $$\begin{aligned} q &\to ar_1 \\ q &\to ar_2 \end{aligned}$$

     kde $q, r_1, r_2, p \in Q$, $a \in \Sigma$, taková, že $r_1 \neq r_2$

   - Z toho je zřejmé, že v gramatice G nemůžou existovat dvě různé posloupnosti přímých derivací začínajících počátečním neterminálem, na jejichž konci by byla tatáž věta. Tím pádem v gramatice G musí pro každou větu, kterou v ní lze vygenerovat, existovat pouze jeden derivační strom s koncovými uzly tvořícími tuto větu. Takto sestavená gramatika G musí tudíž vždy být jednoznačná.
   - Z předchozích bodů plyne, že každý regulární jazyk lze převést na ekvivalentní jednoznačnou gramatiku, a tudíž je dokázáno, že pro každý regulární jazyk existuje jednoznačná gramatika.