

# Cvičenie 3 - Regulárne výrazy

Ing. Viliam Hromada, PhD.

C-510  
Ústav informatiky a matematiky  
FEI STU

`viliam.hromada@stuba.sk`



## Aký jazyk popisuje regulárny výraz?

Slovne popíšte jazyky dané uvedenými regulárnymi výrazmi:

1.  $R_1 = \varepsilon|a|b$
2.  $R_2 = (a|b)(a|b)^*$
3.  $R_3 = 0(0|1)^*1$
4.  $R_4 = (begin|end)$
5.  $R_5 = ((b|B)(e|E)(g|G)(i|I)(n|N))|((e|E)(n|N)(d|D))$
6.  $R_6 = (a|b|\dots|z)^*nos(a|b|\dots|z)^*$ , kde  $(a|b|\dots|z)$  je zjednotenie všetkých malých písmen
7.  $R_7 = (a|b|\dots|z)^*(nos|pin)(a|b|\dots|z)^*$ , kde  $(a|b|\dots|z)$  je zjednotenie všetkých malých písmen



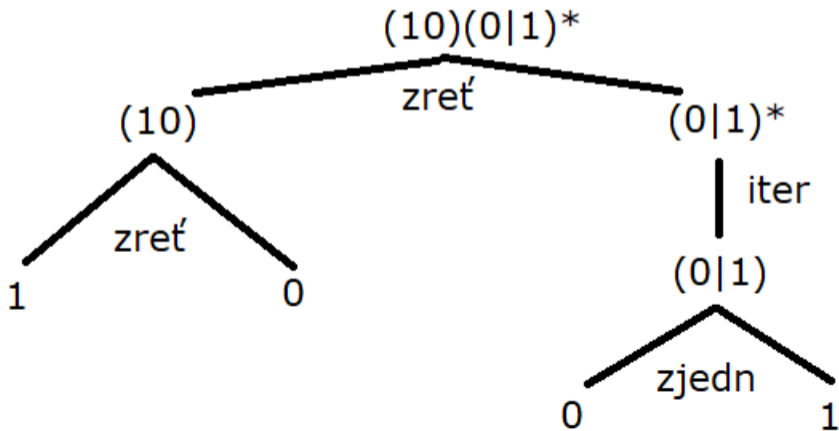
- $R_1$  - jazyk obsahující 3 řetězce  $\varepsilon, a, b$ , t.j. :  $L(R_1) = \{\varepsilon, a, b\}$ .
- $R_2$  - jazyk složený z řetězců nad abecedou  $\{a, b\}$  délky aspoň 1, t.j.  
 $L(R_2) = \{a, b\}^+ = \{w \in \{a, b\}^* \mid |w| \geq 1\} =$   
 $\{aw \mid w \in \{a, b\}^* \cup \{bw \mid w \in \{a, b\}^*\}$
- $R_3$  - jazyk složený z binárních řetězců začínajících nulou a končících jednotkou, t.j.  $L(R_3) = \{0w1 \mid w \in \{0, 1\}^*\}$
- $R_4$  - jazyk složený z 2 řetězců: *begin* a *end*



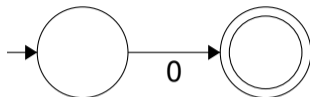
- $R_5$  - jazyk zložený zo všetkých verzií reťazcov *begin* a *end* s malými a veľkými písmenami,  
 $L(R_5) = \{begin, Begin, bEgin, BEgin, beGin, BeGin, \dots, end, \dots, END\}$ .
- $R_6$  - jazyk zložený zo všetkých reťazcov malých písmen, ktoré obsahujú reťazec *nos* ako podreťazec, napr. *nos, podnos, nosom, donosit, ...*
- $R_7$  - jazyk zložený zo všetkých reťazcov malých písmen, ktoré obsahujú alebo reťazec *nos*, alebo reťazec *pin* ako podreťazec, napr. *podnos, camping, aminoskupina, ...*

## Konverzia Regex -> NKA č. 1

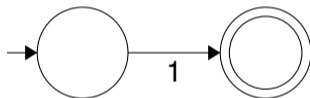
**Príklad:** Zostrojte NKA rozpoznávajúci jazyk popísaný regulárnym výrazom  $((10)(0|1)^*)$ . Rozpíšme si, ako vznikol daný regex zo základných regexov cez operácie zret'azenie, zjednotenie, iterácia:



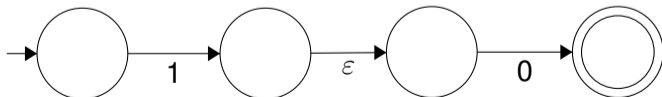
Konečný automat pre regex 0:



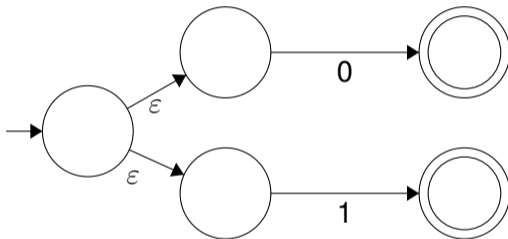
Konečný automat pre regex 1:



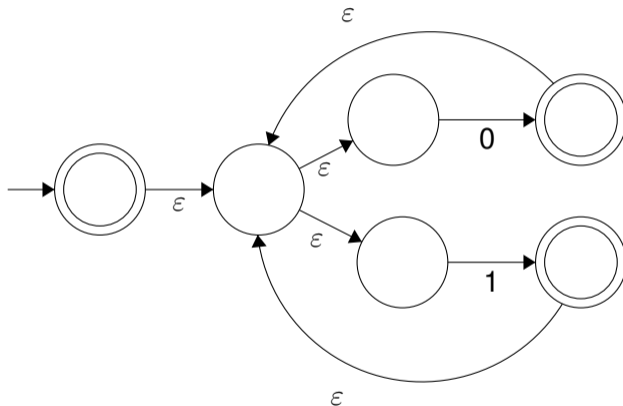
Konečný automat pre regex 10 (t.j. "zreťazenie" 2 už existujúcich automatov).



Konečný automat pre regex  $0 \mid 1$  (t.j. "zjednotenie" 2 už existujúcich automatov).

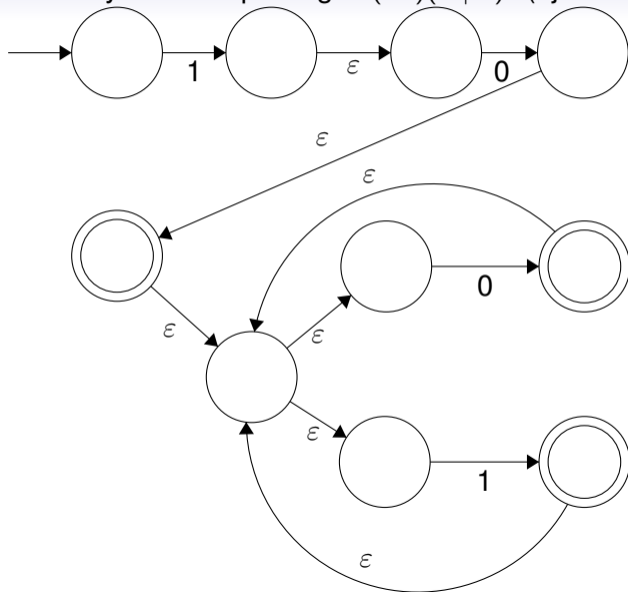


Konečný automat pre regex  $(0 | 1)^*$  (t.j. "iterácia" už existujúceho automatu).



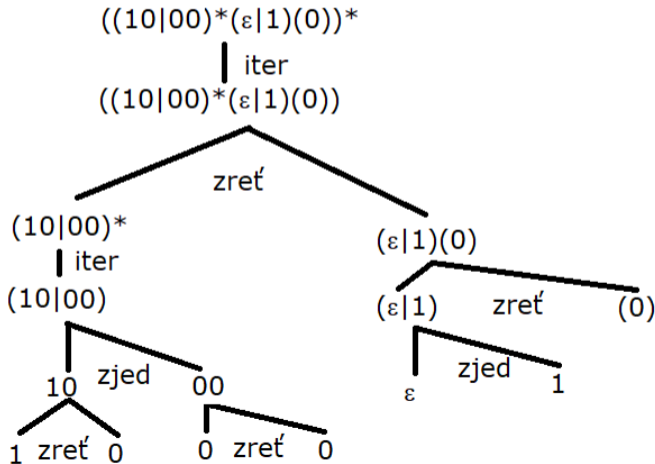


Konečný automat pre regex  $(10)(0 | 1)^*$  (t.j. zreťazenie 10 a  $(0 | 1)^*$ ).

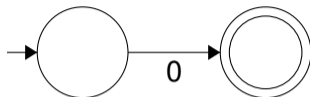


## Konverzia Regex -> NKA č. 2

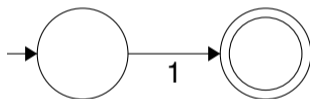
Zostrojte NKA rozpoznávajúci jazyk popísaný regulárnym výrazom:  
 $((10|00)^*(\epsilon|1)(0))^*$ . Znovu si ho najprv rozpišme podľa operácií



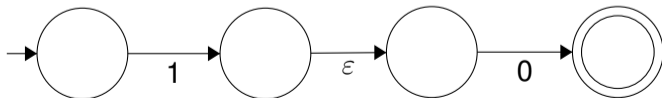
Konečný automat pre regex 0:



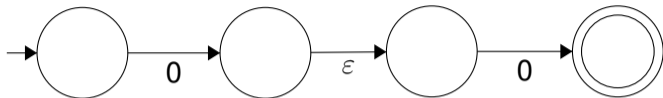
Konečný automat pre regex 1:



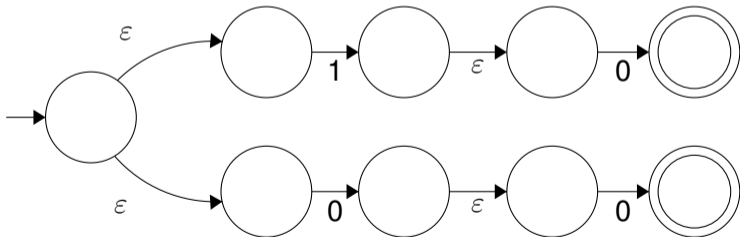
Konečný automat pre regex 10 (t.j. "zreťazenie" 2 už existujúcich automatov).



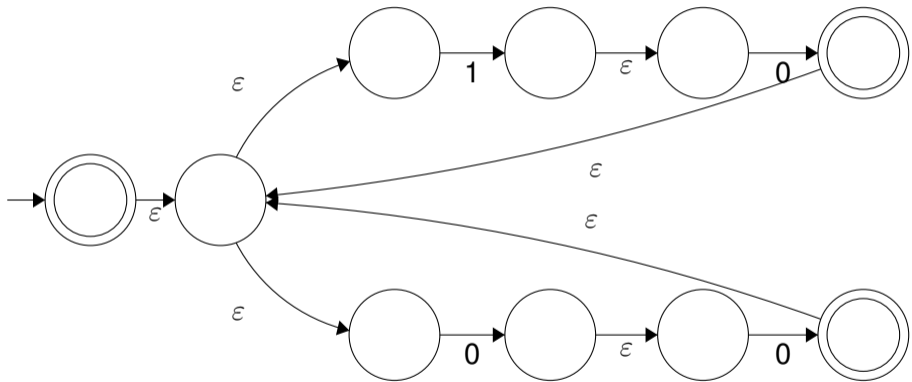
Konečný automat pre regex 00 (t.j. "zreťazenie" 2 už existujúcich automatov).



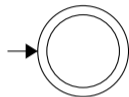
Konečný automat pre regex (10|00) (t.j. "zjednotenie" 2 už existujúcich automatov):



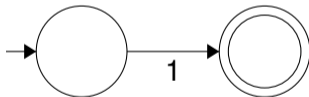
Konečný automat pre regex  $(10|00)^*$  (t.j. "iterácia" už existujúceho automatu):



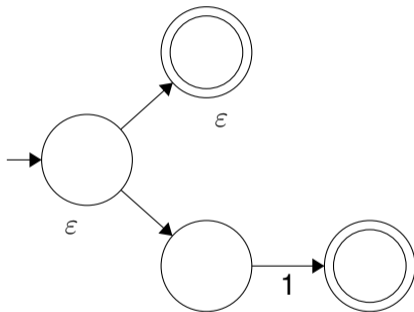
Konečný automat pro regex  $\epsilon$ :



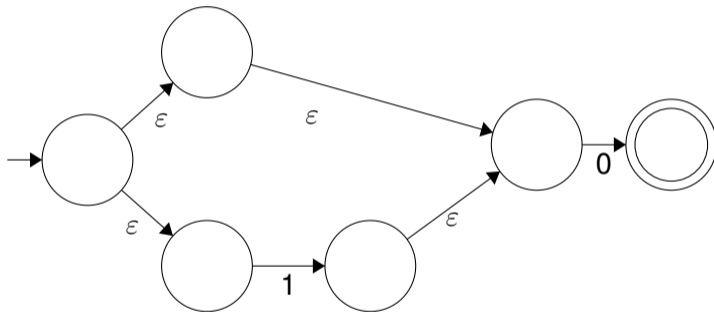
Konečný automat pro regex 1:



Konečný automat pre regex  $\varepsilon|1$  (zjednotenie KA pre  $\varepsilon$  a 1)

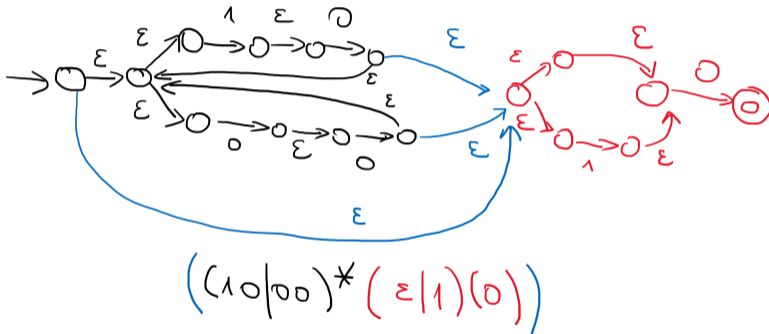


Konečný automat pre regex  $(\epsilon|1)(0)$ :

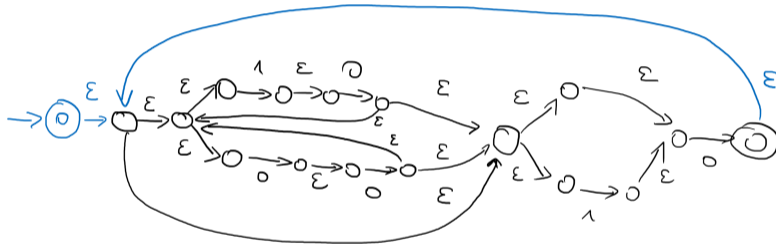




Konečný automat pre regex  $((10|00)^*(\varepsilon|1)(0))$ , ktorý vznikol ako zreťazenie KA pre  $(10|00)^*$  (čiernou farbou) a KA pre  $(\varepsilon|1)(0)$  (červenou farbou). Modrou sú pridané  $\varepsilon$ -prechody.



Konečný automat pre regex  $((10|00)^*(\epsilon|1)(0))^*$ , ktorý vznikol ako iterácia KA z predošlého slajdu. Modrou sú zvýraznené zmeny.



$$\left( (10|00)^*(\epsilon|1)0 \right)^*$$

## Nájdite reprezentáciu uvedeného jazyka

Nájdite regulárnu gramatiku, deterministický konečný automat a regulárny výraz, ktorý popisuje uvedené jazyky.

- $L_1 = \{aw \mid w \in \{a, b\}^*\}$
- $L_2 = \{wa \mid w \in \{a, b\}^*\}$
- $L_3 = \{xaby \mid x \in \{a, b\}^*, y \in \{a, b\}^*\}$
- $L_4 = \{a^*b^*\}$
- $L_5 = \{a, b\}^*$
- $L_6 = \{w \in \{a, b\}^* \mid \#_a(w) \equiv \#_b(w) \pmod{2}\}$
- $L_7 = \{w \in \{a, b\}^* \mid \#_a(w) \equiv \#_b(w) \pmod{3}\}$
- $L_8 = \{w \in \{a, b\}^* \mid \#_a(w) \equiv 1 \pmod{3}\}$
- $L_9 = \{w \in \{a, b\}^* \mid w \text{ má párny počet znakov } b\}$



## Nájdite reprezentáciu uvedeného jazyka

Nájdite regulárnu gramatiku, deterministický konečný automat a regulárny výraz, ktorý popisuje uvedené jazyky.

- $L_{10} = \{w \in \{a, b, c\}^* \mid w \text{ obsahuje ako podreťazec } aba \text{ a začína } c\}$
- $L_{11} = \{a^*b^*\} \cap \{b^*a^*\}$
- $L_{12} = \{a^*b^*\} \cup \{b^*a^*\}$
- $L_{13} = \{b^*ab^*\} \cap \{a^*bab\}$
- $L_{14} = \{w \in \{a, b\}^* \mid \#_a(w) < 3\}$
- $L_{15} = \{w \in \{a, b\}^* \mid \#_a(w) < 3\}^C$  vzhľadom na abecedu  $A = \{a, b\}^*$
- $L_{16} = \{aw \mid w \in \{a, b\}^*\} \setminus \{wa \mid w \in \{a, b\}^*\}$
- $L_{17} = \{w \mid \{1, 2, 3, 4, 5, 6, 7, 8, 9, 0\}^* \mid w \text{ je prirodzené číslo deliteľné } 3\}$

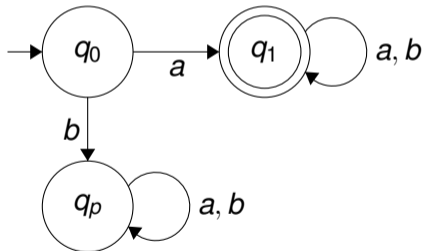


$$L_1 = \{aw \mid w \in \{a, b\}^*\}$$

$G = (N, T, P, S)$ ,  $N = \{S, A\}$ ,  $T = \{a, b\}$ ,  $S$  je počiatkový neterminál, pravidlá  $P$ :

- $S \rightarrow aA$
- $A \rightarrow aA \mid bA \mid \varepsilon$

DKA:



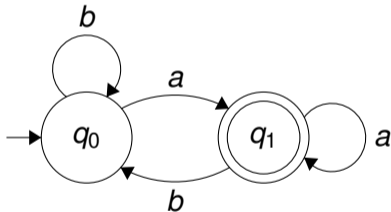
Regex:  $a(a|b)^*$

$$L_2 = \{wa \mid w \in \{a, b\}^*\}$$

$G = (N, T, P, S)$ ,  $N = \{S\}$ ,  $T = \{a, b\}$ ,  $S$  je počiatočný neterminál, pravidlá  $P$ :

- $S \rightarrow aS \mid bS \mid a$

DKA:



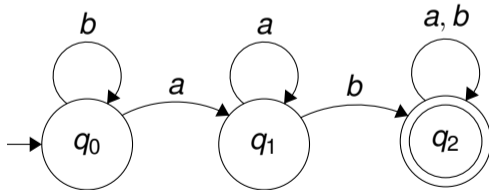
Regex:  $(a|b)^*a$

$$L_3 = \{xaby \mid x \in \{a, b\}^*, y \in \{a, b\}^*\}$$

$G = (N, T, P, S)$ ,  $N = \{S, A, B\}$ ,  $T = \{a, b\}$ ,  $S$  je počiatočný neterminál, pravidlá  $P$ :

- $S \rightarrow aS \mid bS \mid aA$
- $A \rightarrow bB$
- $B \rightarrow aB \mid bB \mid \varepsilon$

DKA:



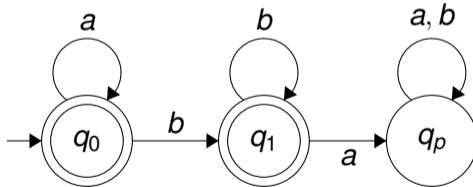
Regex:  $(a|b)^* ab(a|b)^*$

$$L_4 = \{a^*b^*\}$$

$G = (N, T, P, S)$ ,  $N = \{S, A, B\}$ ,  $T = \{a, b\}$ ,  $S$  je počiatočný neterminál, pravidlá  $P$ :

- $S \rightarrow \varepsilon \mid aA \mid bB$
- $A \rightarrow \varepsilon \mid aA \mid bB$
- $B \rightarrow \varepsilon \mid bB$

DKA:



Regex:  $a^*b^*$

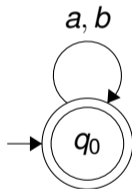


$$L_5 = \{a, b\}^*$$

$G = (N, T, P, S)$ ,  $N = \{S\}$ ,  $T = \{a, b\}$ ,  $S$  je počiatočný neterminál, pravidlá  $P$ :

- $S \rightarrow \varepsilon \mid aS \mid bS$

DKA:



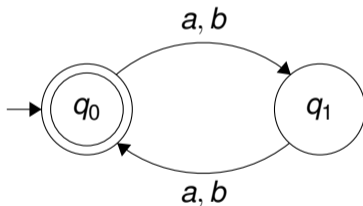
Regex:  $(a|b)^*$

$$L_6 = \{w \in \{a, b\}^* \mid \#_a(w) \equiv \#_b(w) \pmod{2}\}$$

$G = (N, T, P, S)$ ,  $N = \{S, A\}$ ,  $T = \{a, b\}$ ,  $S$  je počiatkový neterminál, pravidlá  $P$ :

- $S \rightarrow \varepsilon \mid aA \mid bA$
- $A \rightarrow aS \mid bS$

DKA:



Regex:  $((a|b)(a|b))^*$

*(Pozn. jazyk tvoria reťazce, v ktorých má počet  $a$  a  $b$  rovnakú paritu - inými slovami reťazce s párnou dĺžkou)*

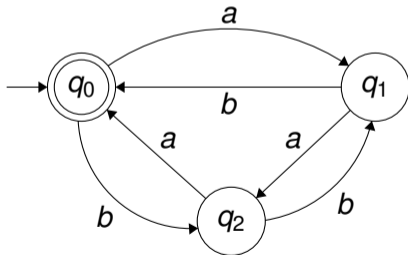


$$L_7 = \{w \in \{a, b\}^* \mid \#_a(w) \equiv \#_b(w) \pmod{3}\}$$

$G = (N, T, P, S)$ ,  $N = \{S, A, B\}$ ,  $T = \{a, b\}$ ,  $S$  je počiatkový neterminál, pravidlá  $P$ :

- $S \rightarrow \varepsilon \mid aA \mid bB$
- $A \rightarrow aB \mid bS$
- $B \rightarrow aS \mid bA$

DKA:



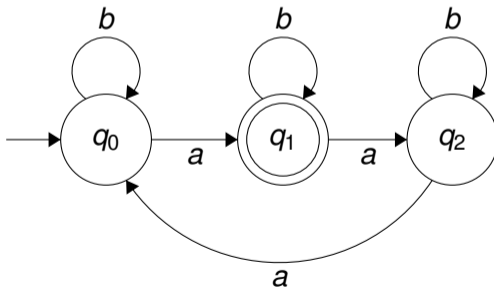
Regex: *nemusíte vedieť urobiť, je zbytočne zložitý...*

$$L_8 = \{w \in \{a, b\}^* \mid \#_a(w) \equiv 1 \pmod{3}\}$$

$G = (N, T, P, S)$ ,  $N = \{S, A, B\}$ ,  $T = \{a, b\}$ ,  $S$  je počiatkový neterminál, pravidlá  $P$ :

- $S \rightarrow bS \mid aA$
- $A \rightarrow \varepsilon \mid bA \mid aB$
- $B \rightarrow aS \mid bB$

DKA:



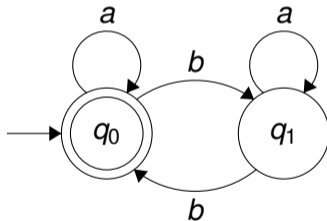
Regex:  $(b^*ab^*ab^*a)^*b^*ab^*$

$$L_9 = \{w \in \{a, b\}^* \mid w \text{ má párny počet znakov } b\}$$

$G = (N, T, P, S)$ ,  $N = \{S, A\}$ ,  $T = \{a, b\}$ ,  $S$  je počiatkový neterminál, pravidlá  $P$ :

- $S \rightarrow \varepsilon \mid bA \mid aS$
- $A \rightarrow \varepsilon aA \mid bS$

DKA:



Regex:  $(a^*)|(a^*ba^*ba^*)^*$ , prípadne  $(a|(ba^*b))^*$ .

$L_{10} = \{w \in \{a, b, c\}^* \mid w \text{ obsahuje ako podreťazec } aba \text{ a začína } c\}$

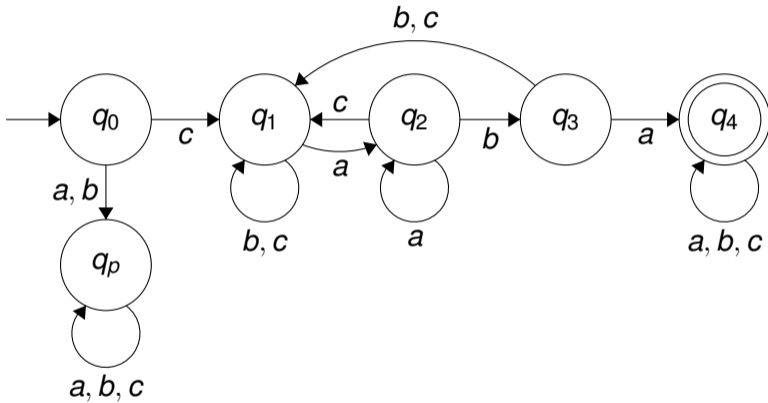
$G = (N, T, P, S)$ ,  $N = \{S, A, B\}$ ,  $T = \{a, b, c\}$ ,  $S$  je počiatočný neterminál,  $P$ :

- $S \rightarrow cA$
- $A \rightarrow aA \mid bA \mid cA \mid abaB$
- $B \rightarrow aB \mid bB \mid cB \mid \varepsilon$

Regex:  $c(a|b|c)^*aba(a|b|c)^*$

DKA na ďalšom slajde.



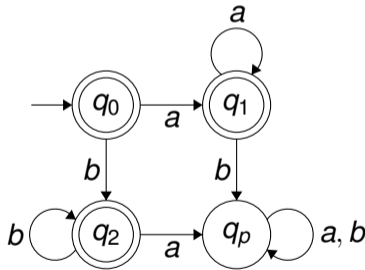


$$L_{11} = \{a^*b^*\} \cap \{b^*a^*\}$$

$G = (N, T, P, S)$ ,  $N = \{S, A, B\}$ ,  $T = \{a, b\}$ ,  $S$  je počiatkový neterminál,  $P$ :

- $S \rightarrow \varepsilon \mid aA \mid bB$
- $A \rightarrow \varepsilon \mid aA$
- $B \rightarrow \varepsilon \mid bB$

DKA:



Regex:  $(a^*)|(b^*)$

Pozn.: V podstate ide alebo o reťazce  $a^*$ , alebo o reťazce  $b^*$

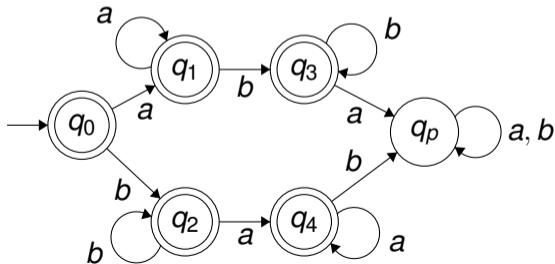


$$L_{12} = \{a^*b^*\} \cup \{b^*a^*\}$$

$G = (N, T, P, S)$ ,  $N = \{S, A, B, C, D\}$ ,  $T = \{a, b\}$ ,  $S$  je počiatkový neterminál,  $P$ :

- $S \rightarrow \varepsilon \mid aA \mid bB \mid bC \mid aD$
- $A \rightarrow \varepsilon \mid aA \mid bB$
- $B \rightarrow \varepsilon \mid bB$
- $C \rightarrow \varepsilon \mid bC \mid aD$
- $D \rightarrow \varepsilon \mid aD$

DKA:



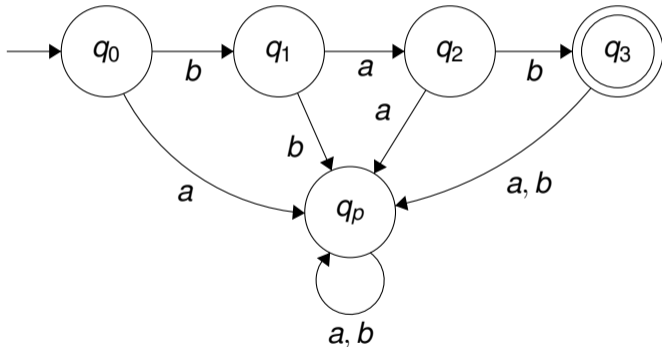
Regex:  $(a^*b^*)|(b^*a^*)$

$$L_{13} = \{b^*ab^*\} \cap \{a^*bab\}$$

$G = (N, T, P, S)$ ,  $N = \{S\}$ ,  $T = \{a, b\}$ ,  $S$  je počiatkový neterminál,  $P$ :

- $S \rightarrow bab$

DKA:



Regex:  $bab$

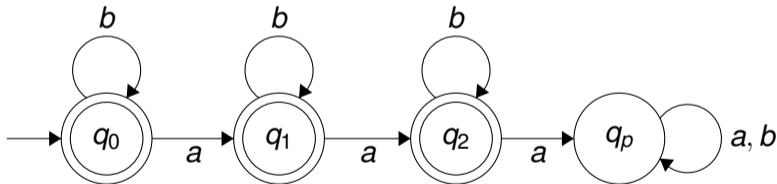
Pozn.: V podstate ide len o reťazec  $bab$

$$L_{14} = \{w \in \{a, b\}^* \mid \#_a(w) < 3\}$$

$G = (N, T, P, S)$ ,  $N = \{S, A, B\}$ ,  $T = \{a, b\}$ ,  $S$  je počiatočný neterminál,  $P$ :

- $S \rightarrow \varepsilon \mid bS \mid aA$
- $A \rightarrow \varepsilon \mid bA \mid aB$
- $B \rightarrow \varepsilon \mid bB$

DKA:



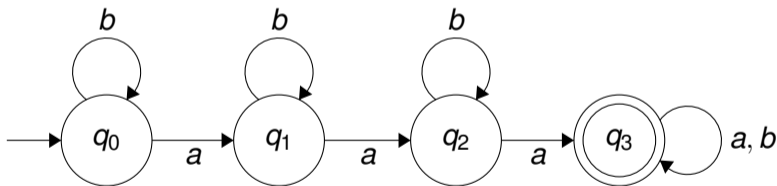
Regex:  $b^*(\varepsilon \mid ab^* \mid ab^* ab^*)$

$$L_{15} = \{w \in \{a, b\}^* \mid \#_a(w) < 3\}^C$$

$G = (N, T, P, S)$ ,  $N = \{S, A, B, C\}$ ,  $T = \{a, b\}$ ,  $S$  je počiatkový neterminál,  $P$ :

- $S \rightarrow bS \mid aA$
- $A \rightarrow bA \mid aB$
- $B \rightarrow bB \mid aC$
- $C \rightarrow \varepsilon \mid bC \mid aC$

DKA:



Regex:  $b^* ab^* ab^* a(a|b)^*$

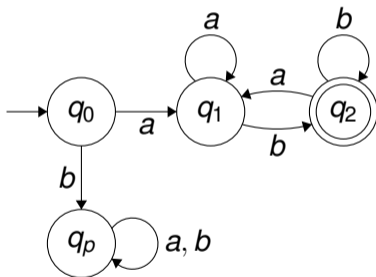
*Pozn.: Ide o doplnok predchádzajúceho jazyka ( $L_{14}$  boli reťazce s najviac 2 áčkami,  $L_{15}$  sú reťazce s aspoň 3 áčkami).*

$$L_{16} = \{aw \mid w \in \{a, b\}^*\} \setminus \{wa \mid w \in \{a, b\}^*\}$$

$G = (N, T, P, S)$ ,  $N = \{S, A, B, C\}$ ,  $T = \{a, b\}$ ,  $S$  je počiatkový neterminál,  $P$ :

- $S \rightarrow aA$
- $A \rightarrow aA \mid bA \mid b$

DKA:



Regex:  $a(a|b)^*b$

(Pozn.: Rozdiel uvedených množín tvorí množina reťazcov začínajúcich  $a$  a končiacich  $b$ )

$$L_{17} = \{w \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 0\}^* \mid w \text{ je prirodzené číslo deliteľné } 3\}$$

$G = (N, T, P, S)$ ,  $N = \{S, A, B, C\}$ ,  $T = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ ,  $S$  je počiatkový neterminál,  $P$ :

- $S \rightarrow 3A \mid 6A \mid 9A \mid 1B \mid 4B \mid 7B \mid 2C \mid 5C \mid 8C$
- $A \rightarrow 0A \mid 3A \mid 6A \mid 9A \mid 1B \mid 4B \mid 7B \mid 2C \mid 5C \mid 8C$
- $B \rightarrow 0B \mid 3B \mid 6B \mid 9B \mid 1C \mid 4C \mid 7C \mid 2A \mid 5A \mid 8A$
- $C \rightarrow 0C \mid 3C \mid 6C \mid 9C \mid 1A \mid 4A \mid 7A \mid 2B \mid 5B \mid 8B$

DKA na ďalšom slajde. Regex nemusíte vedieť, lebo v tomto prípade je dosť zložitý...



