Prednáška 3 - Regulárne výrazy, vlastnosti regulárnych jazykov

Ing. Viliam Hromada, PhD.

C-510
Ústav informatiky a matematiky
FEI STU

viliam.hromada@stuba.sk

KA vs. RegGram

- Množina regulárnych jazykov (t.j. tých, ktoré su generovateľné regulárnymi gramatikami) je rovná množine jazykov rozpoznateľných konečnými automatmi.

Regulárna gramatika $\Rightarrow$ KA

Veta

Nech $G = (N, T, P, S)$ je regulárna gramatika. Potom existuje nedeterministický konečný automat $M = (Q, T, \delta, q_0, F)$ taký, že $L(M) = L(G)$.

Regulárna gramatika $\Rightarrow$ KA

Dôkaz: Výpočet automatu bude kopírovať odvodenie v gramatike.

- Množina stavov Q bude tvorená množinou neterminálov N gramatiky a novým stavom $\{q_f\}$ takým, že $q_f \notin N$ (t.j. $Q = N \cup \{q_f\}$).
- Vstupná abeceda automatu je rovná terminálom gramatiky.
- Začiatočný stav automatu q_0 bude začiatočný neterminál gramatiky, t.j. $q_0 = S$.
- Množina akceptujúcich stavov F:

$$F = \begin{cases} \{S, q_f\} & \text{ak } S \to \varepsilon \in P \\ \{q_f\} & \text{inak.} \end{cases}$$

- Prechodová funkcia δ:
  - ak $B \to a \in P$, potom $q_f \in \delta(B, a)$;
  - ak $B \to aC \in P$, potom $C \in \delta(B, a)$.

Regulárna gramatika $\Rightarrow$ KA

Dôkaz: Výpočet automatu bude kopírovať odvodenie v gramatike.

- Množina stavov Q bude tvorená množinou neterminálov N gramatiky a novým stavom $\{q_f\}$ takým, že $q_f \notin N$ (t.j. $Q = N \cup \{q_f\}$).
- Vstupná abeceda automatu je rovná terminálom gramatiky.
- Začiatočný stav automatu q_0 bude začiatočný neterminál gramatiky, t.j. $q_0 = S$.
- Množina akceptujúcich stavov F:

$$F = \begin{cases} \{S, q_f\} & \text{ak } S \to \varepsilon \in P \\ \{q_f\} & \text{inak.} \end{cases}$$

- Prechodová funkcia δ:
    - ak $B \to a \in P$, potom $q_f \in \delta(B, a)$;
    - ak $B \to aC \in P$, potom $C \in \delta(B, a)$.

Regulárna gramatika $\Rightarrow$ KA

Dôkaz: Výpočet automatu bude kopírovať odvodenie v gramatike.

- Množina stavov Q bude tvorená množinou neterminálov N gramatiky a novým stavom $\{q_f\}$ takým, že $q_f \notin N$ (t.j. $Q = N \cup \{q_f\}$).
- Vstupná abeceda automatu je rovná terminálom gramatiky.
- Začiatočný stav automatu q_0 bude začiatočný neterminál gramatiky, t.j. $q_0 = S$.
- Množina akceptujúcich stavov F:

$$F = \begin{cases} \{S, q_f\} & \text{ak } S \to \varepsilon \in P \\ \{q_f\} & \text{inak.} \end{cases}$$

- Prechodová funkcia δ:
  - ak $B \to a \in P$, potom $q_f \in \delta(B, a)$;
  - ak $B \to aC \in P$, potom $C \in \delta(B, a)$.

Regulárna gramatika $\Rightarrow$ KA

Dôkaz: Výpočet automatu bude kopírovať odvodenie v gramatike.

- Množina stavov Q bude tvorená množinou neterminálov N gramatiky a novým stavom $\{q_f\}$ takým, že $q_f \notin N$ (t.j. $Q = N \cup \{q_f\}$).
- Vstupná abeceda automatu je rovná terminálom gramatiky.
- Začiatočný stav automatu q_0 bude začiatočný neterminál gramatiky, t.j. $q_0 = S$.
- Množina akceptujúcich stavov F:

$$F = \begin{cases} \{S, q_f\} & \text{ak } S \to \varepsilon \in P \\ \{q_f\} & \text{inak.} \end{cases}$$

- Prechodová funkcia δ:
  - ak $B \to a \in P$, potom $q_f \in \delta(B, a)$;
  - ak $B \to aC \in P$, potom $C \in \delta(B, a)$.

Regulárna gramatika $\Rightarrow$ KA

Dôkaz: Výpočet automatu bude kopírovať odvodenie v gramatike.

- Množina stavov Q bude tvorená množinou neterminálov N gramatiky a novým stavom $\{q_f\}$ takým, že $q_f \notin N$ (t.j. $Q = N \cup \{q_f\}$).
- Vstupná abeceda automatu je rovná terminálom gramatiky.
- Začiatočný stav automatu q_0 bude začiatočný neterminál gramatiky, t.j. $q_0 = S$.
- Množina akceptujúcich stavov F:

$$F = \begin{cases} \{S, q_f\} & \text{ak } S \to \varepsilon \in P \\ \{q_f\} & \text{inak.} \end{cases}$$

- Prechodová funkcia δ:
  - ak $B \to a \in P$, potom $q_f \in \delta(B, a)$;
  - ak $B \to aC \in P$, potom $C \in \delta(B, a)$.

Regulárna gramatika $\Rightarrow$ (N)KA

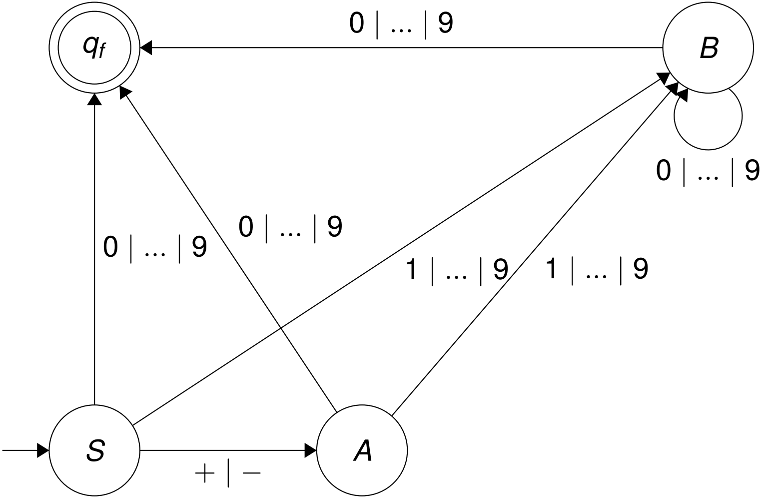
Príklad: Gramatika $G = (\{S, A, B\}, \{+, -, 0, 1, 2, ..., 9\}, P, S)$.

$$S \rightarrow +A \mid -A \mid 1B \mid 2B \mid ... \mid 0 \mid 1 \mid ... \mid 9$$
$$A \rightarrow 1B \mid 2B \mid ... \mid 0 \mid 1 \mid ... \mid 9$$
$$B \rightarrow 0B \mid 1B \mid ... \mid 0 \mid 1 \mid ... \mid 9$$

Hľadaný NKA M:

δ	$+ \mid -$	0	$1 \mid ... \mid 9$
S	$\{A\}$	$\{q_f\}$	$\{B, q_f\}$
A	$\emptyset$	$\{q_f\}$	$\{B, q_f\}$
B	$\emptyset$	$\{B, q_f\}$	$\{B, q_f\}$
q_f	$\emptyset$	$\emptyset$	$\emptyset$

$Q = \{S, A, B, q_f\}$
$T = \{+, -, 0, 1, ..., 9\}$
$q_0 = S$
$F = \{q_f\}$


q_f
B
S
A
0 | ... | 9
0 | ... | 9
0 | ... | 9
0 | ... | 9
1 | ... | 9
1 | ... | 9
+ | −

V gramatike viem odvodiť:

$$S \Rightarrow -A \Rightarrow -1B \Rightarrow -12B \Rightarrow -123$$

t.j. slovo -123 patrí do jazyka generovaného gramatikou, $-123 \in L(G)$, a teda by sa malo dať akceptovať automatom:

$$(S, -123) \vdash (A, 123) \vdash (B, 23) \vdash (B, 3) \vdash (q_f, \varepsilon)$$

t.j. naozaj, $-123 \in L(M)$.

Z NKA vytvoriť DKA je už len triviálna úloha.

(D)KA $\Rightarrow$ regulárna gramatika

Veta

Nech $M = (Q, T, \delta, q_0, F)$ je DKA. Potom existuje taká regulárna gramatika $G = (N, T, P, S)$, že $L(G) = L(M)$.

(D)KA $\Rightarrow$ regulárna gramatika

Dôkaz: Odvodenie v gramatike bude kopírovať výpočet automatu.

- Množina neterminálov gramatiky bude totožná s množinou stavov, t.j. $N = Q$.
- Množina terminálov gramatiky bude totožná so vstupnou abecedou DKA.
- Začiatočný symbol gramatiky bude totožný so začiatočným stavom DKA, t.j. $S = q_0$.
- Množina pravidiel P:
    - ak $\delta(q_i, a) = q_j$, potom $q_i \to aq_j \in P$,
    - ak $\delta(q_i, a) = q_f$, $q_f \in F$, potom $q_i \to a \in P$,
    - ak $q_0 \in F$, potom $q_0 \to \varepsilon \in P$.

(D)KA $\Rightarrow$ regulárna gramatika

Dôkaz: Odvodenie v gramatike bude kopírovať výpočet automatu.

- Množina neterminálov gramatiky bude totožná s množinou stavov, t.j. $N = Q$.
- Množina terminálov gramatiky bude totožná so vstupnou abecedou DKA.
- Začiatočný symbol gramatiky bude totožný so začiatočným stavom DKA, t.j. $S = q_0$.
- Množina pravidiel P:
  - ak $\delta(q_i, a) = q_j$, potom $q_i \to aq_j \in P$,
  - ak $\delta(q_i, a) = q_f, q_f \in F$, potom $q_i \to a \in P$,
  - ak $q_0 \in F$, potom $q_0 \to \varepsilon \in P$.

(D)KA $\Rightarrow$ regulárna gramatika

Dôkaz: Odvodenie v gramatike bude kopírovať výpočet automatu.

- Množina neterminálov gramatiky bude totožná s množinou stavov, t.j. $N = Q$.
- Množina terminálov gramatiky bude totožná so vstupnou abecedou DKA.
- Začiatočný symbol gramatiky bude totožný so začiatočným stavom DKA, t.j. $S = q_0$.
- Množina pravidiel P:
  - ak $\delta(q_i, a) = q_j$, potom $q_i \rightarrow aq_j \in P$,
  - ak $\delta(q_i, a) = q_f$, $q_f \in F$, potom $q_i \rightarrow a \in P$,
  - ak $q_0 \in F$, potom $q_0 \rightarrow \varepsilon \in P$.

(D)KA $\Rightarrow$ regulárna gramatika

Dôkaz: Odvodenie v gramatike bude kopírovať výpočet automatu.

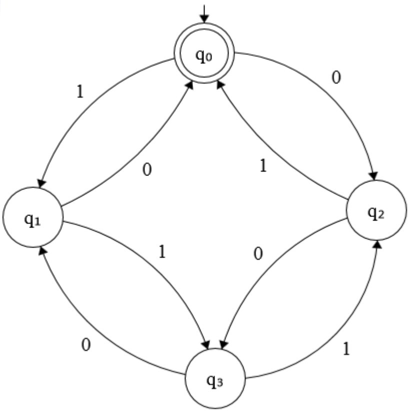
- Množina neterminálov gramatiky bude totožná s množinou stavov, t.j. $N = Q$.
- Množina terminálov gramatiky bude totožná so vstupnou abecedou DKA.
- Začiatočný symbol gramatiky bude totožný so začiatočným stavom DKA, t.j. $S = q_0$.
- Množina pravidiel P:
  - ak $\delta(q_i, a) = q_j$, potom $q_i \to aq_j \in P$,
  - ak $\delta(q_i, a) = q_f, q_f \in F$, potom $q_i \to a \in P$,
  - ak $q_0 \in F$, potom $q_0 \to \varepsilon \in P$.

(D)KA $\Rightarrow$ regulárna gramatika

Dôkaz: Odvodenie v gramatike bude kopírovať výpočet automatu.

- Množina neterminálov gramatiky bude totožná s množinou stavov, t.j. $N = Q$.
- Množina terminálov gramatiky bude totožná so vstupnou abecedou DKA.
- Začiatočný symbol gramatiky bude totožný so začiatočným stavom DKA, t.j. $S = q_0$.
- Množina pravidiel P:
  - ak $\delta(q_i, a) = q_j$, potom $q_i \to aq_j \in P$,
  - ak $\delta(q_i, a) = q_f, q_f \in F$, potom $q_i \to a \in P$,
  - ak $q_0 \in F$, potom $q_0 \to \varepsilon \in P$.

Príklad: Je daný automat (prechodovým diagramom):



Gramatika G:
$N = \{q_0, q_1, q_2, q_3\}$
$T = \{0, 1\}$
$S = q_0$
P obsahuje pravidlá:

- $q_0 \rightarrow \varepsilon$ (lebo $q_0 \in F$)
- $q_0 \rightarrow 0q_2 \mid 1q_1$
- $q_1 \rightarrow 0q_0 \mid 1q_3 \mid 0$ (lebo $q_0 \in F$)
- $q_2 \rightarrow 0q_3 \mid 1q_0 \mid 1$ (lebo $q_0 \in F$)
- $q_3 \rightarrow 0q_1 \mid 1q_2$

Automat napríklad akceptuje slovo 0110:

$$(q_0, 0110) \vdash (q_2, 110) \vdash (q_0, 10) \vdash (q_1, 0) \vdash (q_0, \varepsilon),$$

keďže $q_0 \in F$, tak $0110 \in L(M)$ a teda by ho gramatika mala vedieť vygenerovať:

$$q_0 \Rightarrow 0q_2 \Rightarrow 01q_0 \Rightarrow 011q_1 \Rightarrow 0110,$$

t.j. $0110 \in L(G)$.

Regulárne výrazy

Aby toho nebolo málo, tak regulárne jazyky okrem gramatík a automatov vieme popísať aj inak.

Regulárne výrazy

Definícia

Nech T je abeceda. Potom:

1. *$\emptyset$ je regulárny výraz popisujúci prázdny jazyk,*
2. *ε je regulárny výraz popisujúci jazyk $\{\varepsilon\}$,*
3. *a, pre $a \in T$ je regulárny výraz popisujúci jazyk $\{a\}$.*
4. *Ak R_1 a R_2 sú regulárne výrazy popisujúce jazyky R_1 a R_2, potom $(R_1 \mid R_2)$ je regulárny výraz popisujúci jazyk $R_1 \cup R_2$ (zjednotenie).*
5. *Ak R_1 a R_2 sú regulárne výrazy popisujúce jazyky R_1 a R_2, potom $(R_1 R_2)$ je regulárny výraz popisujúci jazyk $R_1 R_2$ (zreťazenie).*
6. *Ak R je regulárny výraz popisujúci jazyk R, potom (R^*) je regulárny výraz popisujúci iteráciu R^* jazyka R.*
7. *iné regulárne výrazy ako tie, zostrojené podľa bodov 1-6, neexistujú.*

Regulárne výrazy

Definícia

Nech T je abeceda. Potom:

1. *$\emptyset$ je regulárny výraz popisujúci prázdny jazyk,*
2. *ε je regulárny výraz popisujúci jazyk $\{\varepsilon\}$,*
3. *a, pre $a \in T$ je regulárny výraz popisujúci jazyk $\{a\}$.*
4. *Ak R_1 a R_2 sú regulárne výrazy popisujúce jazyky R_1 a R_2, potom $(R_1 \mid R_2)$ je regulárny výraz popisujúci jazyk $R_1 \cup R_2$ (zjednotenie).*
5. *Ak R_1 a R_2 sú regulárne výrazy popisujúce jazyky R_1 a R_2, potom $(R_1 R_2)$ je regulárny výraz popisujúci jazyk $R_1 R_2$ (zreťazenie).*
6. *Ak R je regulárny výraz popisujúci jazyk R, potom (R^*) je regulárny výraz popisujúci iteráciu R^* jazyka R.*
7. *iné regulárne výrazy ako tie, zostrojené podľa bodov 1-6, neexistujú.*

Regulárne výrazy

Definícia

Nech T je abeceda. Potom:

1. $\emptyset$ *je regulárny výraz popisujúci prázdny jazyk,*
2. ε *je regulárny výraz popisujúci jazyk* $\{\varepsilon\}$,
3. *a, pre* $a \in T$ *je regulárny výraz popisujúci jazyk* $\{a\}$.
4. *Ak* R_1 *a* R_2 *sú regulárne výrazy popisujúce jazyky* R_1 *a* R_2, *potom* $(R_1 \mid R_2)$ *je regulárny výraz popisujúci jazyk* $R_1 \cup R_2$ *(zjednotenie).*
5. *Ak* R_1 *a* R_2 *sú regulárne výrazy popisujúce jazyky* R_1 *a* R_2, *potom* $(R_1 R_2)$ *je regulárny výraz popisujúci jazyk* $R_1 R_2$ *(zreťazenie).*
6. *Ak* R *je regulárny výraz popisujúci jazyk* R, *potom* (R^*) *je regulárny výraz popisujúci iteráciu* R^* *jazyka* R.
7. *iné regulárne výrazy ako tie, zostrojené podľa bodov 1-6, neexistujú.*

Regulárne výrazy

Definícia

Nech T je abeceda. Potom:

1. *$\emptyset$ je regulárny výraz popisujúci prázdny jazyk,*
2. *ε je regulárny výraz popisujúci jazyk $\{\varepsilon\}$,*
3. *a, pre $a \in T$ je regulárny výraz popisujúci jazyk $\{a\}$.*
4. *Ak R_1 a R_2 sú regulárne výrazy popisujúce jazyky R_1 a R_2, potom $(R_1 \mid R_2)$ je regulárny výraz popisujúci jazyk $R_1 \cup R_2$ (zjednotenie).*
5. *Ak R_1 a R_2 sú regulárne výrazy popisujúce jazyky R_1 a R_2, potom $(R_1 R_2)$ je regulárny výraz popisujúci jazyk $R_1 R_2$ (zreťazenie).*
6. *Ak R je regulárny výraz popisujúci jazyk R, potom (R^*) je regulárny výraz popisujúci iteráciu R^* jazyka R.*
7. *iné regulárne výrazy ako tie, zostrojené podľa bodov 1-6, neexistujú.*

Regulárne výrazy

Definícia

Nech T je abeceda. Potom:

1. $\emptyset$ *je regulárny výraz popisujúci prázdny jazyk,*
2. ε *je regulárny výraz popisujúci jazyk* $\{\varepsilon\}$,
3. *a, pre* $a \in T$ *je regulárny výraz popisujúci jazyk* $\{a\}$.
4. *Ak* R_1 *a* R_2 *sú regulárne výrazy popisujúce jazyky* R_1 *a* R_2, *potom* $(R_1 \mid R_2)$ *je regulárny výraz popisujúci jazyk* $R_1 \cup R_2$ *(zjednotenie).*
5. *Ak* R_1 *a* R_2 *sú regulárne výrazy popisujúce jazyky* R_1 *a* R_2, *potom* $(R_1 R_2)$ *je regulárny výraz popisujúci jazyk* $R_1 R_2$ *(zreťazenie).*
6. *Ak* R *je regulárny výraz popisujúci jazyk* R, *potom* (R^*) *je regulárny výraz popisujúci iteráciu* R^* *jazyka* R.
7. *iné regulárne výrazy ako tie, zostrojené podľa bodov 1-6, neexistujú.*

Regulárne výrazy

Definícia

Nech T je abeceda. Potom:

1. *$\emptyset$ je regulárny výraz popisujúci prázdny jazyk,*
2. *ε je regulárny výraz popisujúci jazyk $\{\varepsilon\}$,*
3. *a, pre $a \in T$ je regulárny výraz popisujúci jazyk $\{a\}$.*
4. *Ak R_1 a R_2 sú regulárne výrazy popisujúce jazyky R_1 a R_2, potom $(R_1 \mid R_2)$ je regulárny výraz popisujúci jazyk $R_1 \cup R_2$ (zjednotenie).*
5. *Ak R_1 a R_2 sú regulárne výrazy popisujúce jazyky R_1 a R_2, potom $(R_1 R_2)$ je regulárny výraz popisujúci jazyk $R_1 R_2$ (zreťazenie).*
6. *Ak R je regulárny výraz popisujúci jazyk R, potom (R^*) je regulárny výraz popisujúci iteráciu R^* jazyka R.*
7. *iné regulárne výrazy ako tie, zostrojené podľa bodov 1-6, neexistujú.*

Regulárne výrazy

Definícia

Nech T je abeceda. Potom:

1. *$\emptyset$ je regulárny výraz popisujúci prázdny jazyk,*
2. *ε je regulárny výraz popisujúci jazyk $\{\varepsilon\}$,*
3. *a, pre $a \in T$ je regulárny výraz popisujúci jazyk $\{a\}$.*
4. *Ak R_1 a R_2 sú regulárne výrazy popisujúce jazyky R_1 a R_2, potom $(R_1 \mid R_2)$ je regulárny výraz popisujúci jazyk $R_1 \cup R_2$ (zjednotenie).*
5. *Ak R_1 a R_2 sú regulárne výrazy popisujúce jazyky R_1 a R_2, potom $(R_1 R_2)$ je regulárny výraz popisujúci jazyk $R_1 R_2$ (zreťazenie).*
6. *Ak R je regulárny výraz popisujúci jazyk R, potom (R^*) je regulárny výraz popisujúci iteráciu R^* jazyka R.*
7. *iné regulárne výrazy ako tie, zostrojené podľa bodov 1-6, neexistujú.*

Regulárne výrazy

Definícia

Nech T je abeceda. Potom:

1. $\emptyset$ *je regulárny výraz popisujúci prázdny jazyk,*
2. ε *je regulárny výraz popisujúci jazyk* $\{\varepsilon\}$,
3. *a, pre* $a \in T$ *je regulárny výraz popisujúci jazyk* $\{a\}$.
4. *Ak* R_1 *a* R_2 *sú regulárne výrazy popisujúce jazyky* R_1 *a* R_2, *potom* $(R_1 \mid R_2)$ *je regulárny výraz popisujúci jazyk* $R_1 \cup R_2$ *(zjednotenie).*
5. *Ak* R_1 *a* R_2 *sú regulárne výrazy popisujúce jazyky* R_1 *a* R_2, *potom* $(R_1 R_2)$ *je regulárny výraz popisujúci jazyk* $R_1 R_2$ *(zreťazenie).*
6. *Ak* R *je regulárny výraz popisujúci jazyk* R, *potom* (R^*) *je regulárny výraz popisujúci iteráciu* R^* *jazyka* R.
7. *iné regulárne výrazy ako tie, zostrojené podľa bodov 1-6, neexistujú.*

Príklad: Nech $T = \{0, 1, ..., 9, a, b, ..., z\}$. Potom možné regulárne výrazy:

- $(0 \mid 1) = \{0, 1\}$
- $(1(0^*)1) = \{11, 101, 1001, ...\}$
- $((10)(0 \mid 1)^*) = \{10w \mid w \in \{0, 1\}^*\}$
- $((begin) \mid (end)) = \{begin, end\}$
- $((0 \mid 1 \mid ... \mid 9)(0 \mid 1 \mid ... \mid 9)^*)$ - celočíselné konštanty s prípadnými bezvýznamnými nulami zľava, bez znamienka
- $((\varepsilon \mid + \mid -)(0 \mid ((1 \mid ... \mid 9)(0 \mid 1 \mid ... \mid 9)^*)))$ - celočíselné konštanty bez/so znamienkom, bez bezvýznamných núl zľava
- $((a \mid ... \mid z)^*(begin)(a \mid ... \mid z)^*)$ - všetky textové reťazce obsahujúce *begin* ako podreťazec

Príklad: Nech $T = \{0, 1, ..., 9, a, b, ..., z\}$. Potom možné regulárne výrazy:

- $(0 \mid 1) = \{0, 1\}$
- $(1(0^*)1) = \{11, 101, 1001, ...\}$
- $((10)(0 \mid 1)^*) = \{10w \mid w \in \{0, 1\}^*\}$
- $((begin) \mid (end)) = \{begin, end\}$
- $((0 \mid 1 \mid ... \mid 9)(0 \mid 1 \mid ... \mid 9)^*)$ - celočíselné konštanty s prípadnými bezvýznamnými nulami zľava, bez znamienka
- $((\varepsilon \mid + \mid -)(0 \mid ((1 \mid ... \mid 9)(0 \mid 1 \mid ... \mid 9)^*)))$ - celočíselné konštanty bez/so znamienkom, bez bezvýznamných núl zľava
- $((a \mid ... \mid z)^*(begin)(a \mid ... \mid z)^*)$ - všetky textové reťazce obsahujúce *begin* ako podreťazec

Príklad: Nech $T = \{0, 1, ..., 9, a, b, ..., z\}$. Potom možné regulárne výrazy:

- $(0 \mid 1) = \{0, 1\}$
- $(1(0^*)1) = \{11, 101, 1001, ...\}$
- $((10)(0 \mid 1)^*) = \{10w \mid w \in \{0, 1\}^*\}$
- $((begin) \mid (end)) = \{begin, end\}$
- $((0 \mid 1 \mid ... \mid 9)(0 \mid 1 \mid ... \mid 9)^*)$ - celočíselné konštanty s prípadnými bezvýznamnými nulami zľava, bez znamienka
- $((\varepsilon \mid + \mid -)(0 \mid ((1 \mid ... \mid 9)(0 \mid 1 \mid ... \mid 9)^*)))$ - celočíselné konštanty bez/so znamienkom, bez bezvýznamných núl zľava
- $((a \mid ... \mid z)^*(begin)(a \mid ... \mid z)^*)$ - všetky textové reťazce obsahujúce *begin* ako podreťazec

Príklad: Nech $T = \{0, 1, ..., 9, a, b, ..., z\}$. Potom možné regulárne výrazy:

- $(0 \mid 1) = \{0, 1\}$
- $(1(0^*)1) = \{11, 101, 1001, ...\}$
- $((10)(0 \mid 1)^*) = \{10w \mid w \in \{0, 1\}^*\}$
- $((begin) \mid (end)) = \{begin, end\}$
- $((0 \mid 1 \mid ... \mid 9)(0 \mid 1 \mid ... \mid 9)^*)$ - celočíselné konštanty s prípadnými bezvýznamnými nulami zľava, bez znamienka
- $((\varepsilon \mid + \mid -)(0 \mid ((1 \mid ... \mid 9)(0 \mid 1 \mid ... \mid 9)^*)))$ - celočíselné konštanty bez/so znamienkom, bez bezvýznamných núl zľava
- $((a \mid ... \mid z)^*(begin)(a \mid ... \mid z)^*)$ - všetky textové reťazce obsahujúce *begin* ako podreťazec

Príklad: Nech $T = \{0, 1, ..., 9, a, b, ..., z\}$. Potom možné regulárne výrazy:

- $(0 \mid 1) = \{0, 1\}$
- $(1(0^*)1) = \{11, 101, 1001, ...\}$
- $((10)(0 \mid 1)^*) = \{10w \mid w \in \{0, 1\}^*\}$
- $((begin) \mid (end)) = \{begin, end\}$
- $((0 \mid 1 \mid ... \mid 9)(0 \mid 1 \mid ... \mid 9)^*)$ - celočíselné konštanty s prípadnými bezvýznamnými nulami zľava, bez znamienka
- $((\varepsilon \mid + \mid -)(0 \mid ((1 \mid ... \mid 9)(0 \mid 1 \mid ... \mid 9)^*)))$ - celočíselné konštanty bez/so znamienkom, bez bezvýznamných núl zľava
- $((a \mid ... \mid z)^*(begin)(a \mid ... \mid z)^*)$ - všetky textové reťazce obsahujúce *begin* ako podreťazec

Príklad: Nech $T = \{0, 1, ..., 9, a, b, ..., z\}$. Potom možné regulárne výrazy:

- $(0 \mid 1) = \{0, 1\}$
- $(1(0^*)1) = \{11, 101, 1001, ...\}$
- $((10)(0 \mid 1)^*) = \{10w \mid w \in \{0, 1\}^*\}$
- $((begin) \mid (end)) = \{begin, end\}$
- $((0 \mid 1 \mid ... \mid 9)(0 \mid 1 \mid ... \mid 9)^*)$ - celočíselné konštanty s prípadnými bezvýznamnými nulami zľava, bez znamienka
- $((\varepsilon \mid + \mid -)(0 \mid ((1 \mid ... \mid 9)(0 \mid 1 \mid ... \mid 9)^*)))$ - celočíselné konštanty bez/so znamienkom, bez bezvýznamných núl zľava
- $((a \mid ... \mid z)^*(begin)(a \mid ... \mid z)^*)$ - všetky textové reťazce obsahujúce *begin* ako podreťazec

Príklad: Nech $T = \{0, 1, ..., 9, a, b, ..., z\}$. Potom možné regulárne výrazy:

- $(0 \mid 1) = \{0, 1\}$
- $(1(0^*)1) = \{11, 101, 1001, ...\}$
- $((10)(0 \mid 1)^*) = \{10w \mid w \in \{0, 1\}^*\}$
- $((begin) \mid (end)) = \{begin, end\}$
- $((0 \mid 1 \mid ... \mid 9)(0 \mid 1 \mid ... \mid 9)^*)$ - celočíselné konštanty s prípadnými bezvýznamnými nulami zľava, bez znamienka
- $((\varepsilon \mid + \mid -)(0 \mid ((1 \mid ... \mid 9)(0 \mid 1 \mid ... \mid 9)^*)))$ - celočíselné konštanty bez/so znamienkom, bez bezvýznamných núl zľava
- $((a \mid ... \mid z)^*(begin)(a \mid ... \mid z)^*)$ - všetky textové reťazce obsahujúce *begin* ako podreťazec

Regulárne výrazy $\Rightarrow$ NKA

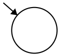
Veta

Nech R je regulárny výraz popisujúci nejaký jazyk. Potom existuje nedeterministický konečný automat M taký, že $L(M) = R$.

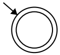
Dôkaz - $\emptyset$, ε, a

Stačí ukázať, že pre prvých 6 bodov z definície vieme vždy zostrojiť príslušný automat:

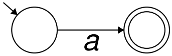
- Výraz $\emptyset$:



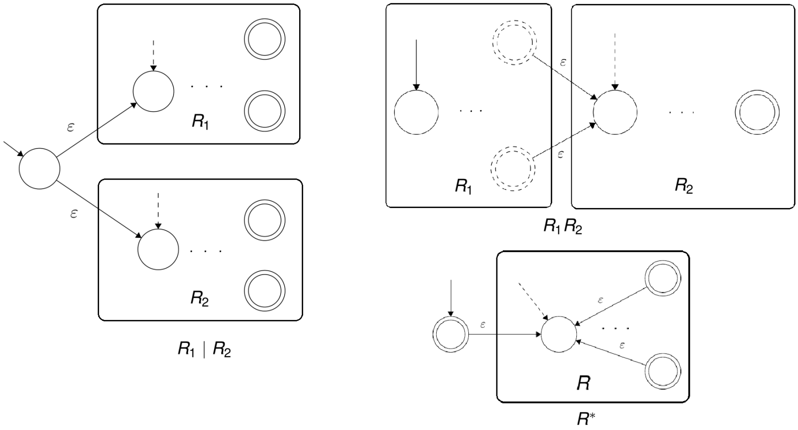
- Výraz ε:



- Výraz a:



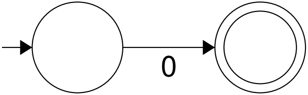
Dôkaz - $R_1 \mid R_2$, $R_1 R_2$, R^*



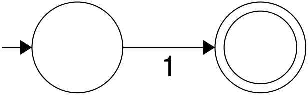
Príklad: Zostrojte NKA rozpoznávajúci jazyk popísaný regulárnym výrazom $((10)(0 \mid 1)^*)$.

Postupne budeme konštruovať konečné automaty podľa konštrukcií z predchádzajúcich slajdov.

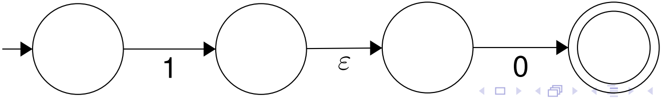
Konečný automat pre jazyk 0:



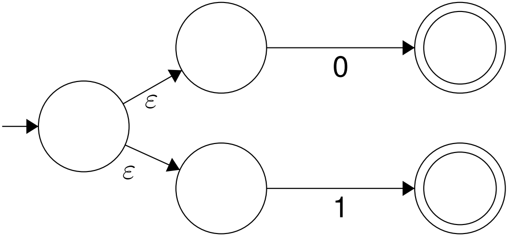
Konečný automat pre jazyk 1:



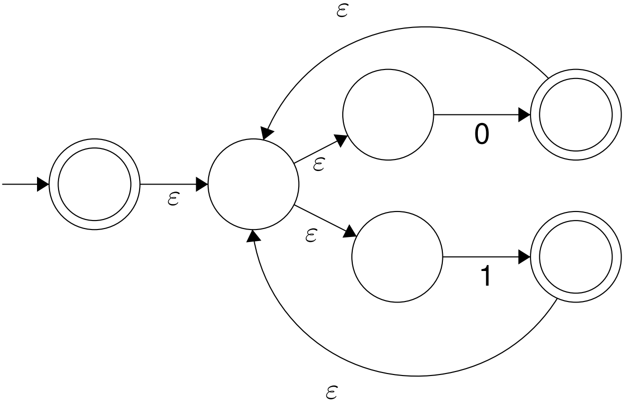
Konečný automat pre jazyk 10 (t.j. "zreťazenie" 2 už existujúcich automatov).



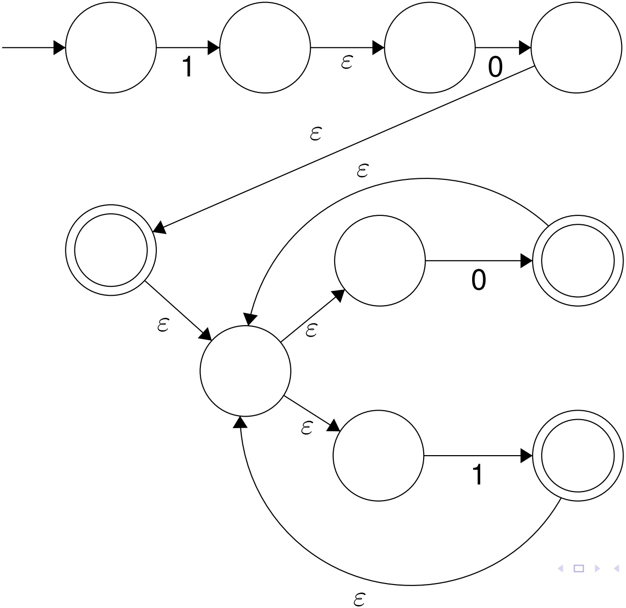
Konečný automat pre jazyk 0 | 1 (t.j. "zjednotenie" 2 už existujúcich automatov).



Konečný automat pre jazyk $(0 \mid 1)^*$ (t.j. "iterácia" už existujúceho automatu).

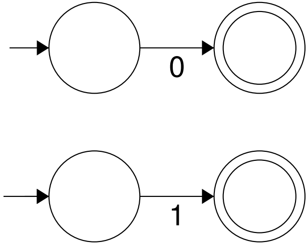


Konečný automat pre jazyk $(10)(0 \mid 1)^*$ (t.j. zreťazenie automatov pre 10 a $(0 \mid 1)^*$.

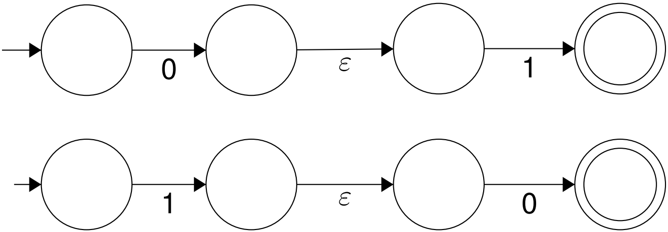


Ďalší príklad

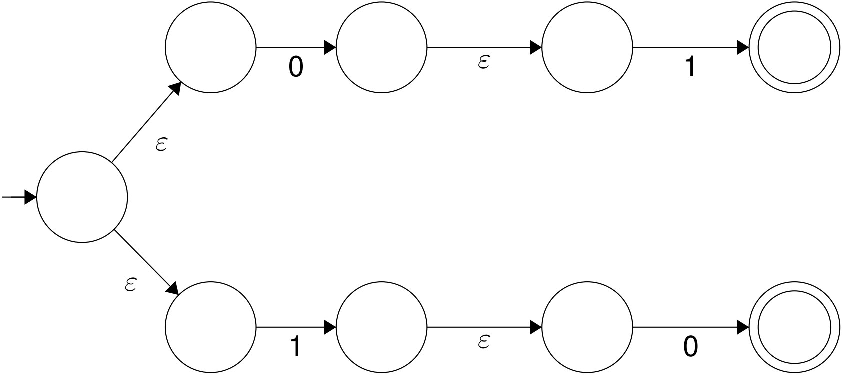
Nájdite minimálny DKA, ktorý akceptuje jazyk popísaný regulárnym výrazom: $(01|10)^*(\varepsilon|0|1)$. Najprv skonštruujeme NKA, ktorý rozpoznáva daný regulárny výraz... Začneme s elementárnymi KA pre 0 a 1.



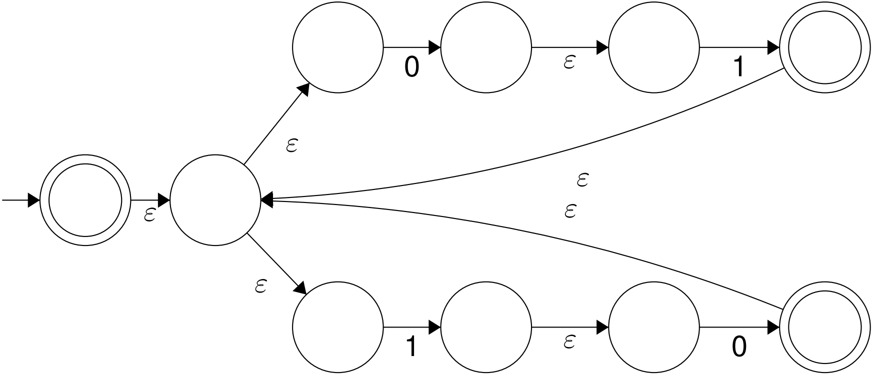
Z nich zreťazením dostávame 2 NKA - prvý pre 01 a druhý pre 10:



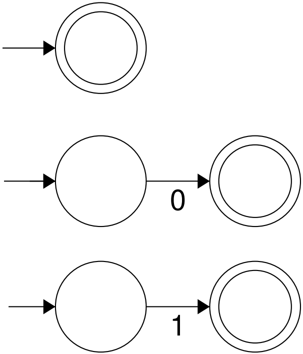
Zjednotením NKA pre 01 a NKA pre 10 vznikne NKA pre $(01|10)$:



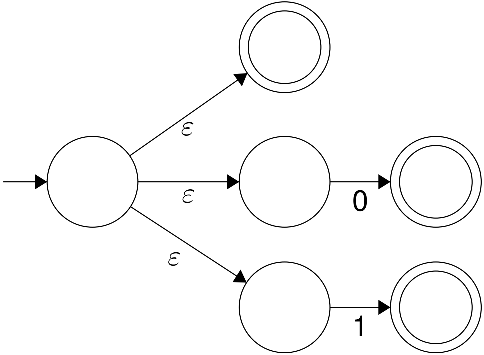
Iteráciou NKA pre (01|10) vznikne NKA pre $(01|10)^*$:



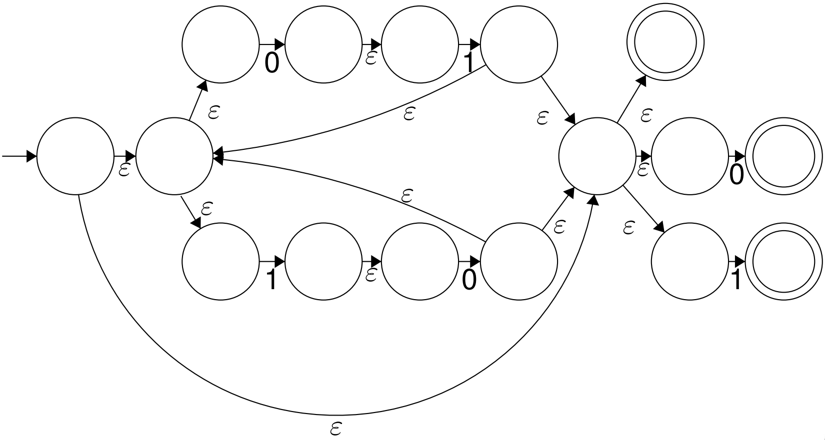
Ďalej potrebujeme zostrojiť NKA pre $(\varepsilon|0|1)$. Tri elementárne NKA, pre $\varepsilon, 0, 1$ sú nasledovné:



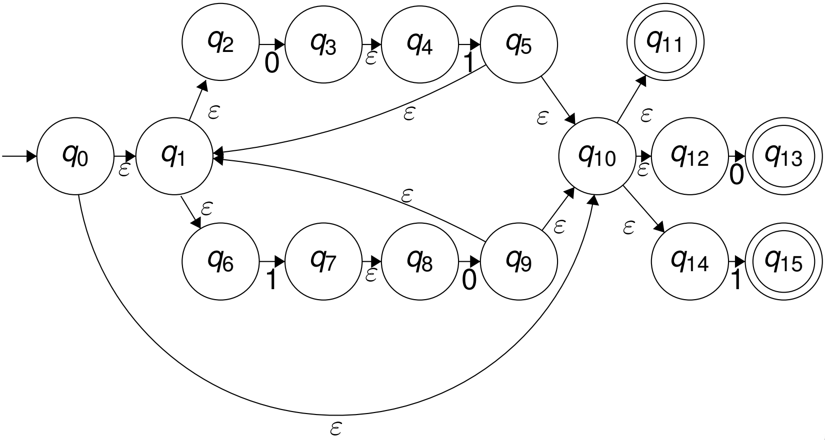
Ich zjednotením dostávame NKA pre $\varepsilon|0|1$:



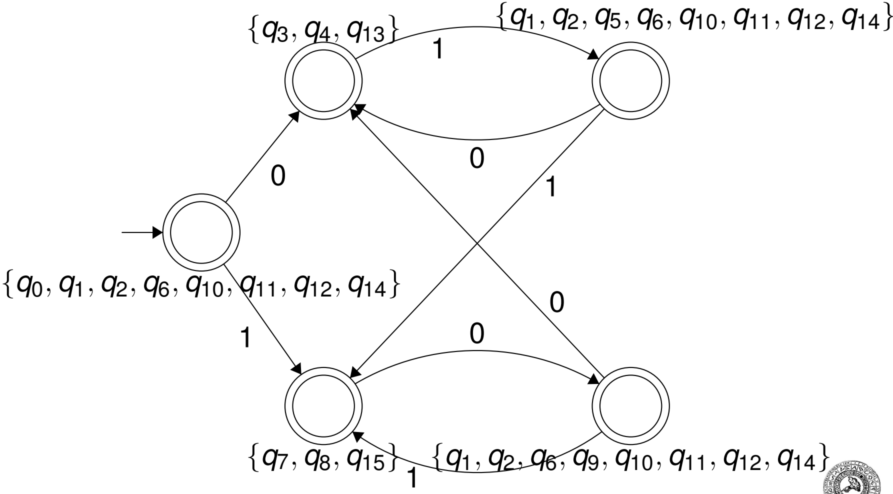
Na záver z NKA pre $(01|10)^*$ a z NKA pre $\varepsilon|0|1$ ich zreťazením dostávame výsledný NKA pre $(01|10)^*(\varepsilon|0|1)$



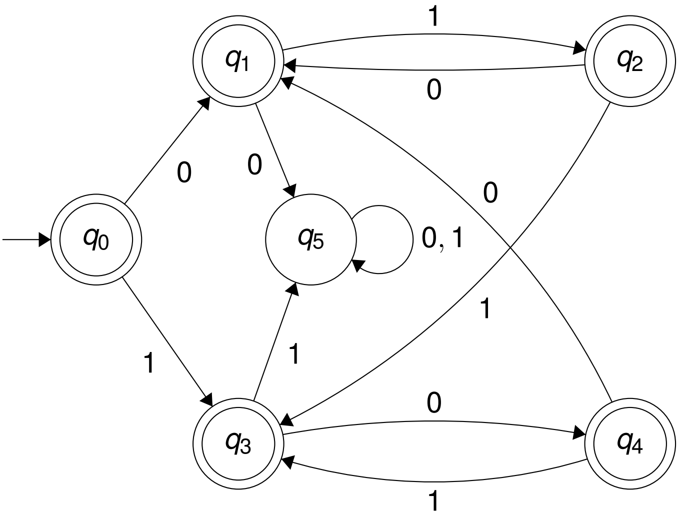
Následne automat podrobíme determinizácii. Najprv však pomenujme stavy (teraz je v princípe jedno, ako).



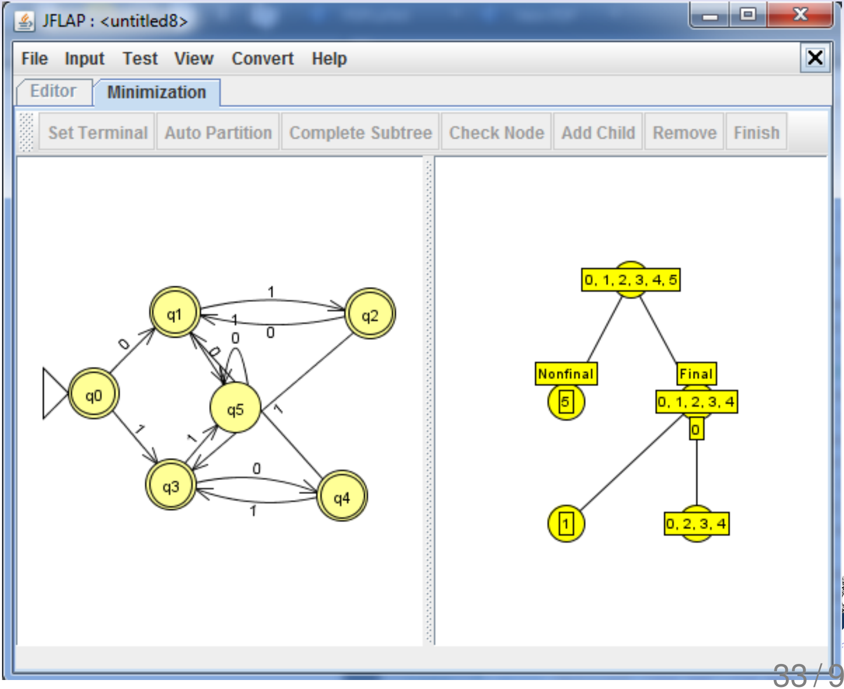
Ekvivalentný deterministický automat vyzerá nasledovne:

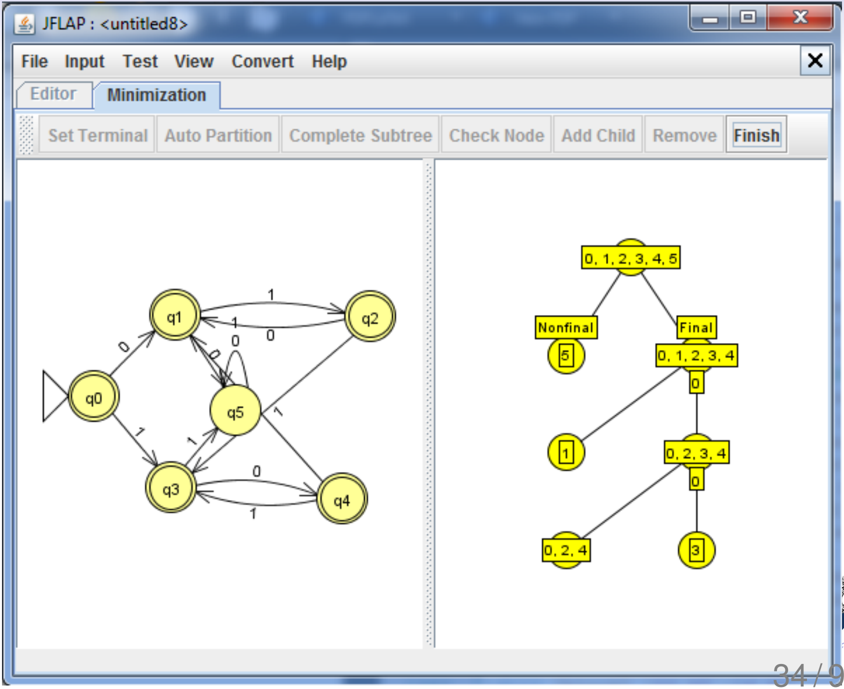


Po preznačení stavov a doplnení na úplný automat (pridáme pascu - stav q_5), dostávame:

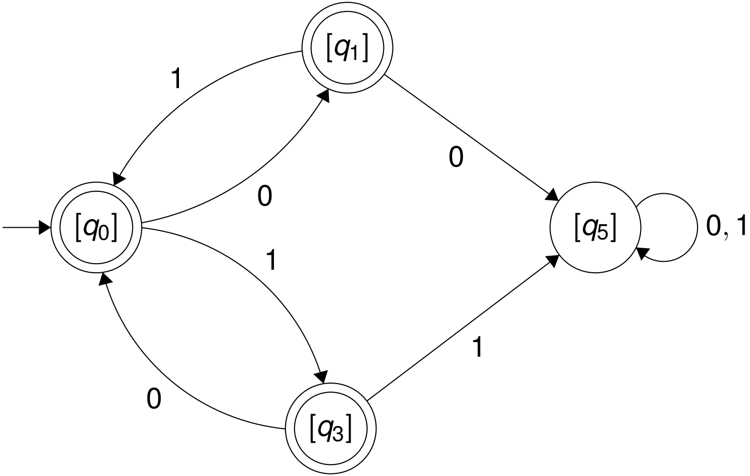


Následne automat podrobíme minimalizácii a dostaneme tak najmenší DKA akceptujúci jazyk popísaný regulárnym výrazom $(01|10)^*(\varepsilon|0|1)$.


JFLAP : <untitled8>
File Input Test View Convert Help
Editor Minimization
Set Terminal Auto Partition Complete Subtree Check Node Add Child Remove Finish
q0
q1
q2
q3
q4
q5
0, 1, 2, 3, 4, 5
Nonfinal
5
Final
0, 1, 2, 3, 4
0
1
0, 2, 3, 4


JFLAP : <untitled8>
File Input Test View Convert Help
Editor
Minimization
Set Terminal
Auto Partition
Complete Subtree
Check Node
Add Child
Remove
Finish
q0
q1
q2
q3
q4
q5
0
1
0, 1, 2, 3, 4, 5
Nonfinal
Final
5
0, 1, 2, 3, 4
0
1
0, 2, 3, 4
0
0, 2, 4
3

A teda výsledný úplný minimálny DKA rozpoznávajúci jazyk popísaný regulárnym výrazom $(01|10)^*(\varepsilon|0|1)$ je:



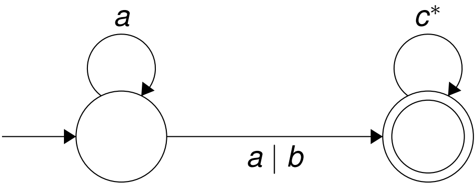
DKA $\Rightarrow$ regulárny výraz

Veta

Nech M je deterministický konečný automat. Potom existuje regulárny výraz R popisujúci taký jazyk, že $L(M) = R$.

Neformálny dôkaz

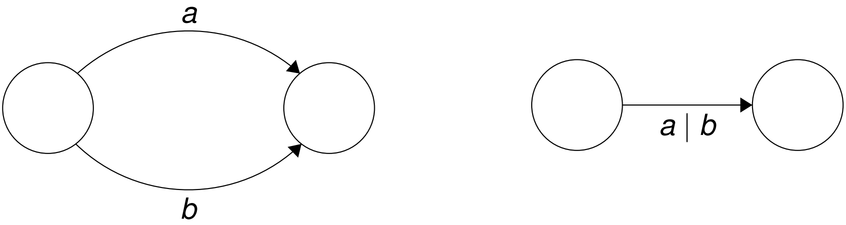
- Grafická reprezentácia DKA/NKA sa niekedy nazýva aj **prechodový diagram**.
- V tomto dôkaze využijeme tzv. **zovšeobecnený prechodový diagram**. Je to prechodový diagram, ktorého hrany sú ohodnotené **regulárnymi výrazmi**, nie len jednoduchými symbolmi.



$$L = a^*(a|b)c^*$$

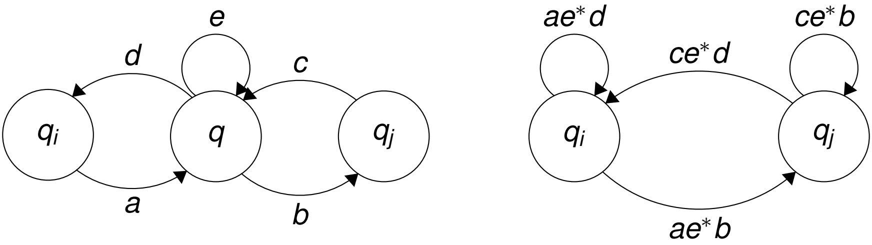
Neformálny dôkaz (pokr.)

Hrany v zovšeobecnenom prechodovom diagrame, ktoré majú spoločný začiatok a spoločný koniec, možno nahradiť jednou hranou, ktorej regulárny výraz bude zjednotením regulárnych výrazov pôvodných hrán.



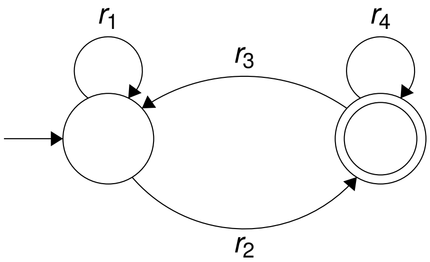
Neformálny dôkaz (pokr.)

Odstránenie stavu q zo zovšeobecneného prechodového diagramu, pričom sa zachová jazyk.



Neformálny dôkaz (pokr.)

Ak bol zovšeobecnený prechodový diagram vytvorený z DKA s 1 akceptačným stavom a odstránime všetky stavy, ktoré nie sú počiatočné alebo akceptačné, dostaneme nasledovnú situáciu (kde r_1, r_2, r_3, r_4 sú nejaké regulárne výrazy):



$$L = r_1^* r_2 (r_4 \mid (r_3 r_1^* r_2))^*$$

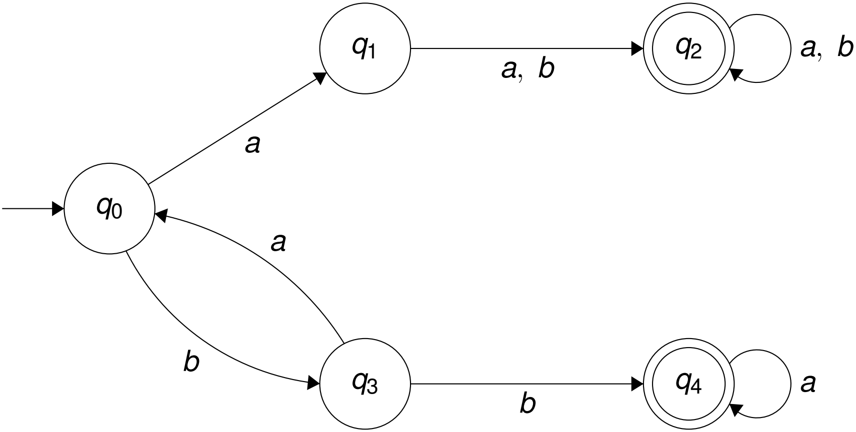
Neformálny dôkaz (pokr.)

1. Je daný DKA s 1 akceptačným stavom.
2. Vytvoríme z neho zovšeobecnený prechodový diagram. Hrany so spoločným začiatkom a spoločným koncom nahradíme novou hranou (viď. slajd č. 38).
3. Odstránime neakceptačné a nepočiatočné stavy podľa slajdu č.39.
4. Intepretujeme jazyk výsledného zovšeobecneného prechodového diagramu s 2 stavmi podľa vzoru zo slajdu č. 40.

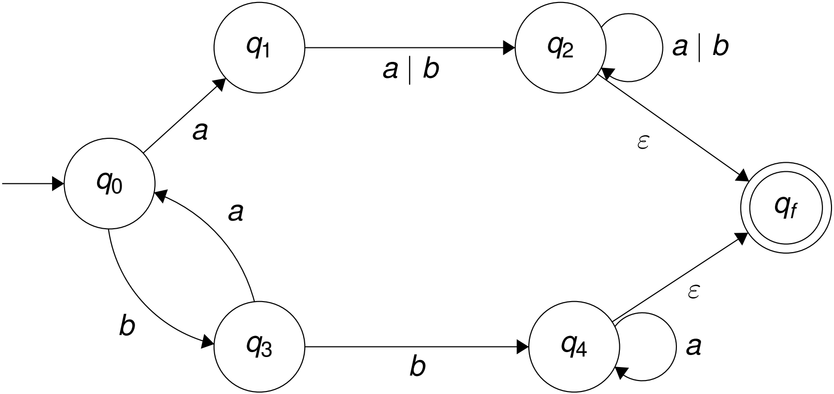
Neformálny dôkaz (pokr.)

- Ak má pôvodný DKA **viacero akceptačných stavov**, tak do prechodového diagramu doplníme nový akceptačný stav, pričom do neho z každého pôvodného akceptačného stavu vedieme hranu ohodnotenú regulárnym výrazom ε.
- Zároveň každý pôvodný akceptačný stav zmeníme na neakceptačný stav.
- Podobne, ak je v pôvodnom DKA **počiatočný** stav q_0 zároveň aj **akceptačným**, tak do prechodového diagramu doplníme nový akceptačný stav a zo stavu q_0 doň vedieme prechod na ε. Zároveň zmeníme q_0 na neakceptačný.
- V konečnom dôsledku dostávame teda popis pôvodného DKA pomocou **regulárneho výrazu**.

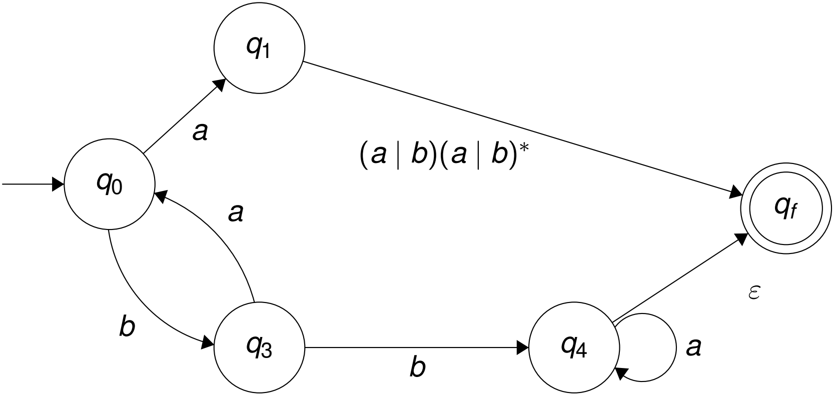
Príklad: Popíšte DKA pomocou regulárneho výrazu



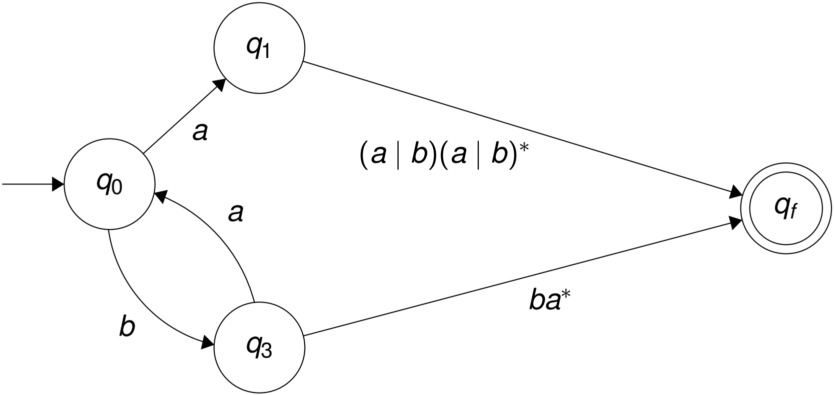
Zovšeobecnený prechodový diagram (doplnený o nový akceptačný stav q_f).



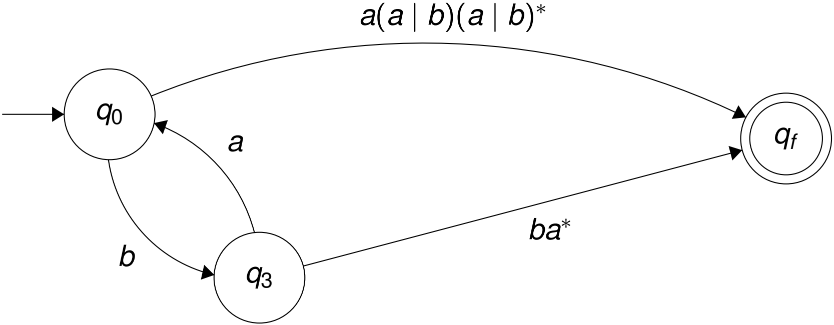
Odstránime q_2:



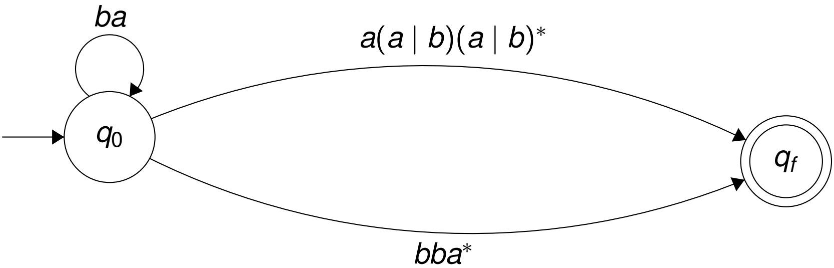
Odstránime q_4:



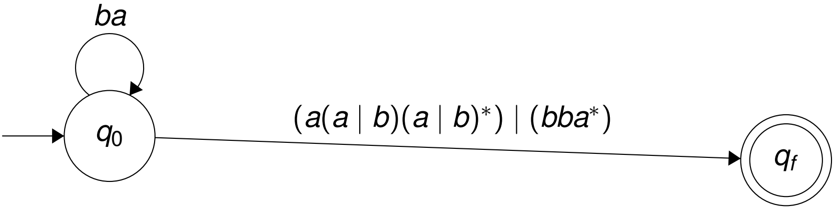
Odstránime q_1:



Odstránime q_3:



Znovu zjednotíme hrany, ktoré majú spoločný začiatok a spoločný koniec:



Výsledný jazyk sa teda dá popísať regulárnym výrazom:

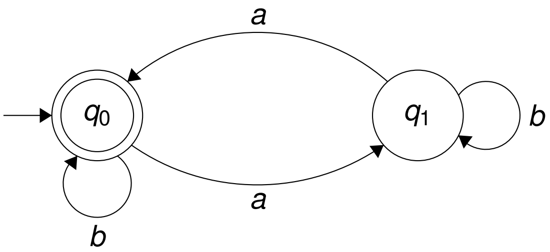
$$(ba)^*((a(a|b)(a|b)^*)|(bba^*))$$

Iné poradie odstraňovania stavov môže viesť k inému tvaru regulárneho výrazu. Jazyk je však stále ten istý!

Príklad na párny počet písmen *a*

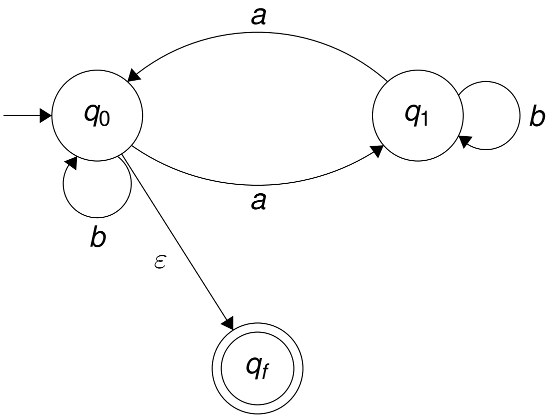
Akým regulárnym jazykom sa dá popísať jazyk $L = \{w \mid w \in \{a,b\}^*, \sharp_a(w) \equiv 0 \bmod 2\}$? $\sharp_a(w)$ znamená počet písmen a v slove w.

Ak príslušný DKA je:

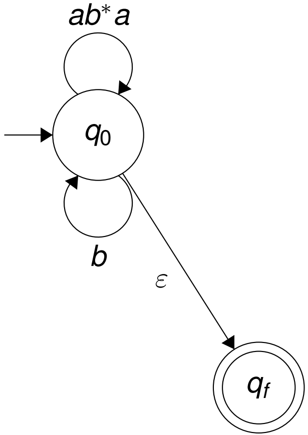


Tak ak z neho urobíme regulárny výraz, nájdeme odpoveď na našu otázku...

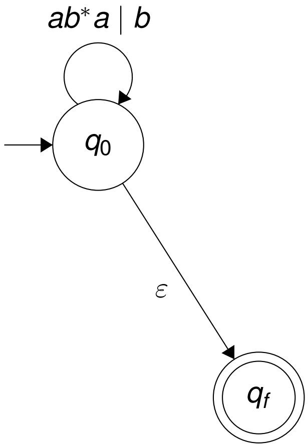
Podľa pokynov najprv pridáme nový akceptačný stav q_f a z q_0 doň vedieme ε-hranu a zároveň z q_0 spravíme neakceptačný stav.



Následne odstránime stav q_1 podľa pokynov...



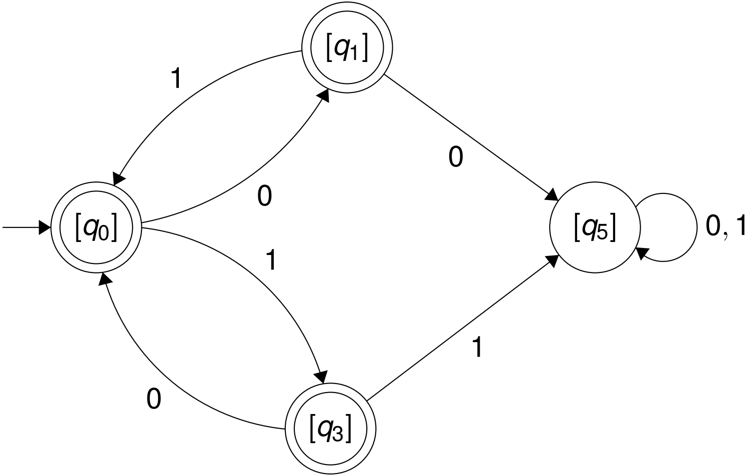
Zjednotíme slučky na stave q_0:



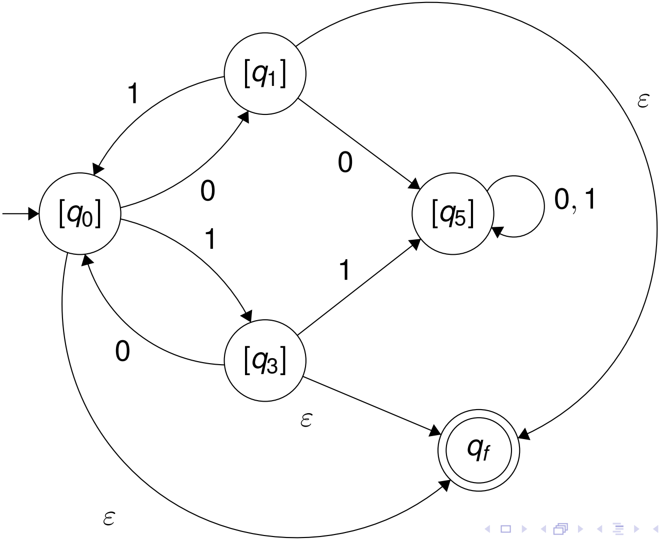
A teda výsledný regulárny výraz je: $((ab^*a)|b)^*\varepsilon$.

Ďalší príklad

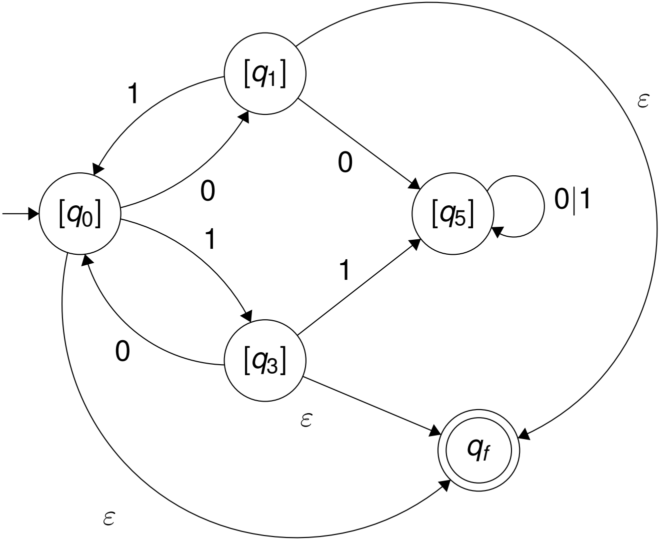
Vezmime DKA zo slajdu č. 35 a nájdime podľa uvedeného postupu ekvivalentný regulárny výraz:



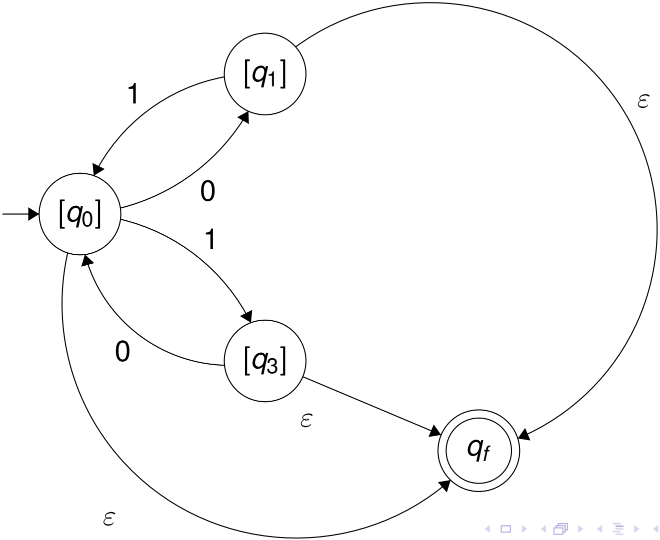
Najprv každý akceptačný stav označíme ako neakceptačný a vedieme ε-hranu do nového akceptačného stavu q_f .



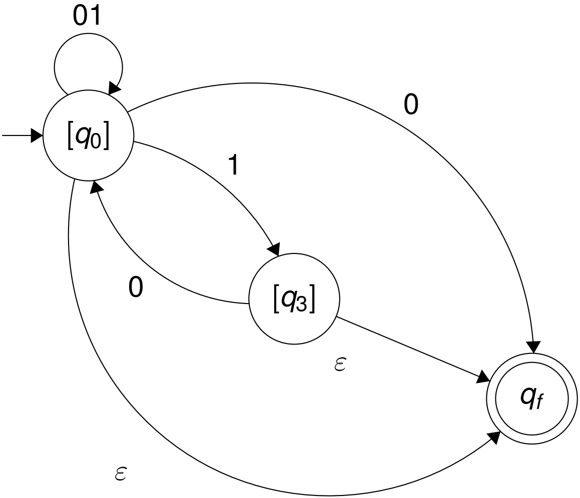
Slučku v $[q_5]$ upravíme, aby obsahovala regulárny výraz.



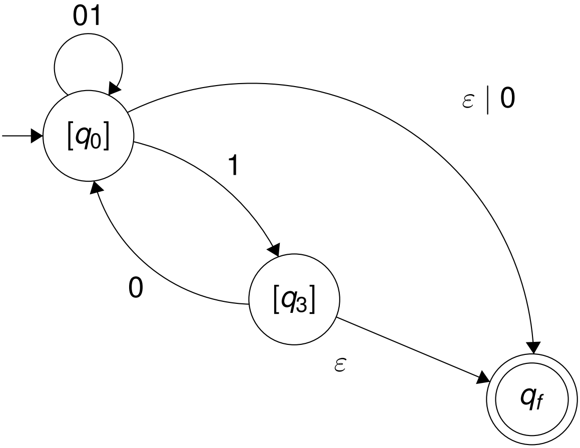
Odstránime stav $[q_5]$. Keďže to bola pasca, do automatu nepribudnú žiadne hrany označené regulárnym výrazom.



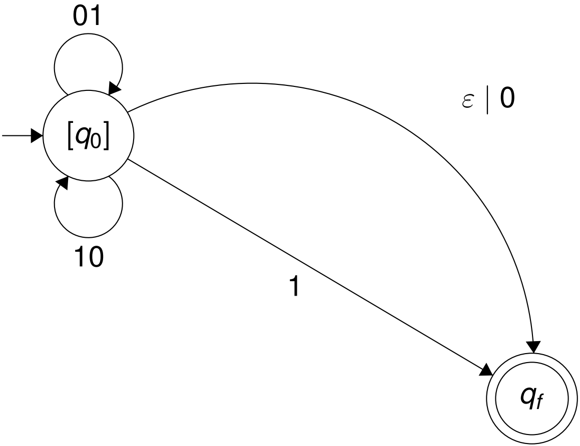
Odstránime stav $[q_1]$. Pribudne 01 slučka v stave $[q_0]$ a hrana 0 z $[q_0]$ do q_f.



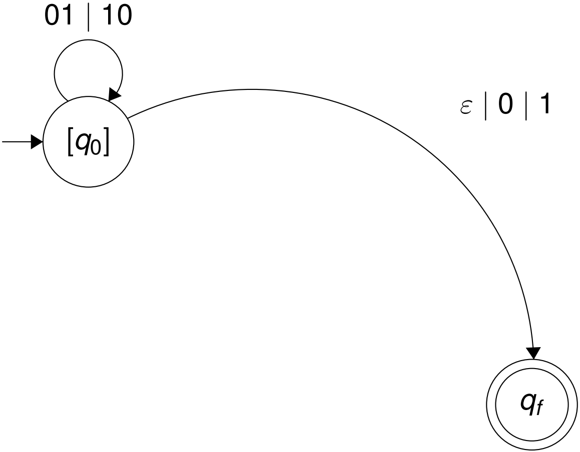
Zjednotíme hrany 0 a ε z $[q_0]$ do q_f.



Odstránime stav $[q_3]$, čím vznikne slučka 10 v $[q_0]$ a hrana 1 z $[q_0]$ do q_f.



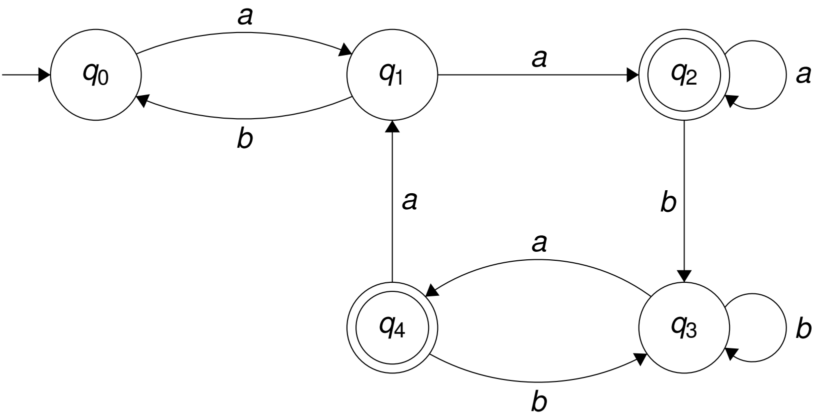
Zjednotíme slučky v $[q_0]$ a hrany z $[q_0]$ do q_f.



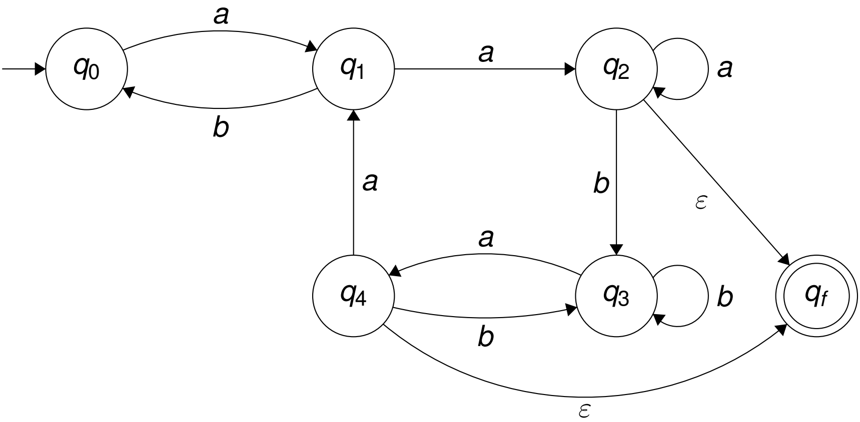
Podľa slajdu č. 40 teda pôvodný DKA akceptuje jazyk: $(01|10)^*(\varepsilon|0|1)$.

A posledný...

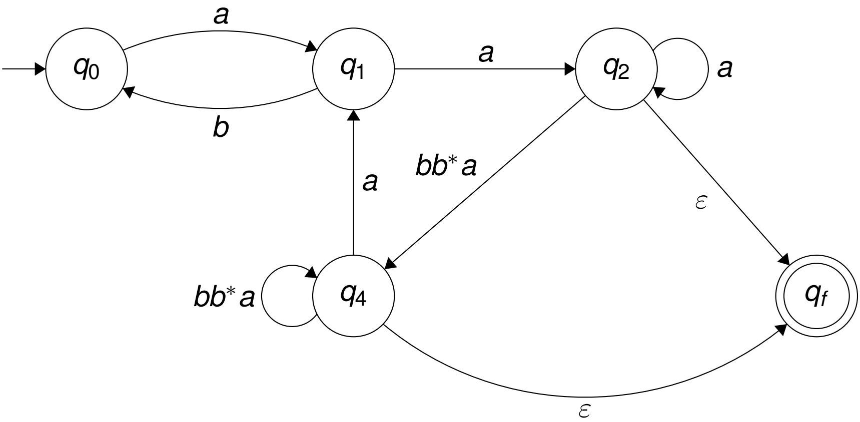
Nájdite regulárny výraz pre DKA:



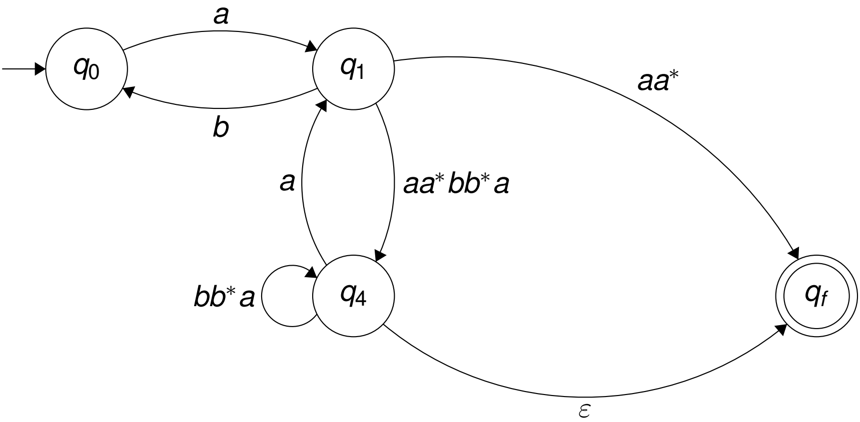
Pridáme q_f...



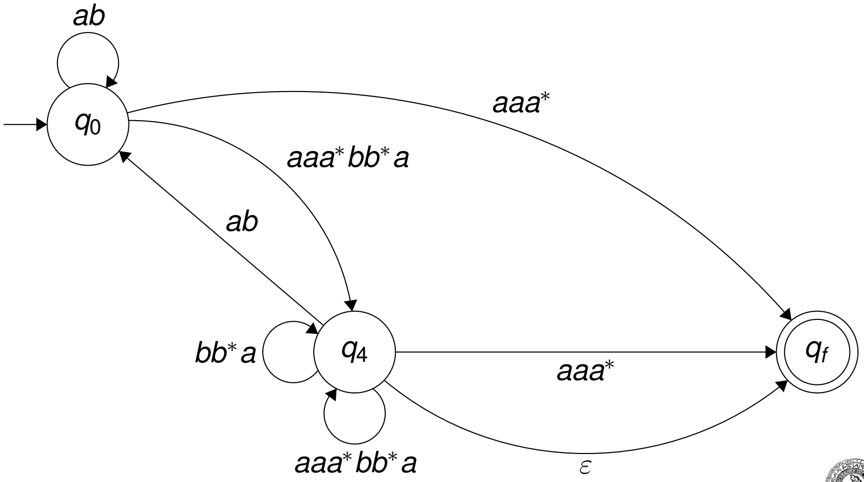
Odstránime q_3...



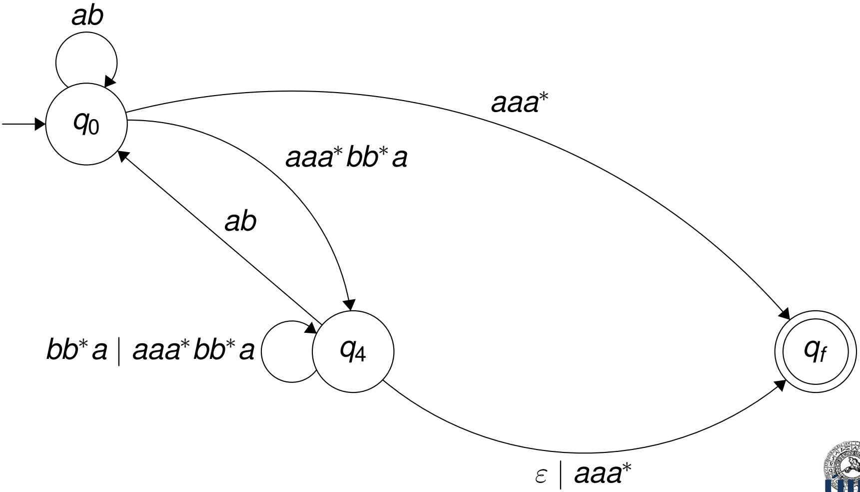
Odstránime q_2...



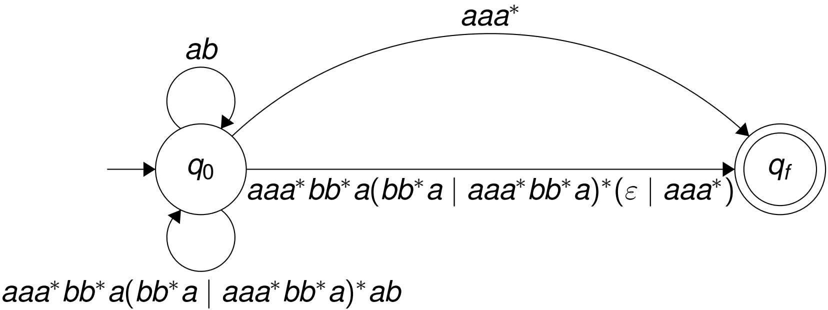
Odstránime q_1...



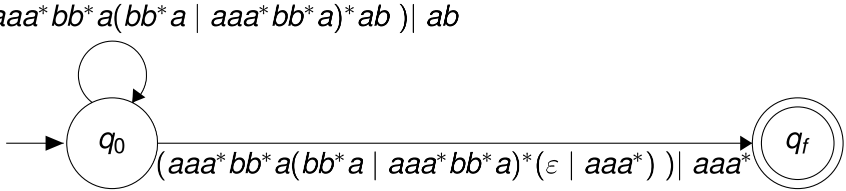
Zjednotíme hrany so spoločným začiatkom a spoločným koncom...



Odstránime q_4:



Zjednotíme hrany so spoločným začiatkom a spoločným koncom...



A teda výsledný regulárny výraz:

$((aaa^*bb^*a(bb^*a \mid aaa^*bb^*a)^*ab) \mid ab)^*((aaa^*bb^*a(bb^*a \mid aaa^*bb^*a)^*(\varepsilon \mid aaa^*)) \mid aaa^*)$

Zhrnutie

Ukázali sme si, ako sa dá:

- Regulárna gramatika previesť na DKA.
- DKA previesť na regulárnu gramatiku.
- Regulárny výraz previesť na NKA (a ten na DKA).
- Previesť DKA na regulárny výraz.

Záver teda je, že:

- Jazyky, ktoré vieme **akceptovať konečnými automatmi** sú presne tie, ktoré vieme **popísať regulárnymi gramatikami** a zároveň presne tie, ktoré vieme **popísať regulárnymi výrazmi**.

Vlastnosti regulárnych jazykov

- Regulárne jazyky sa teda dajú popísať:
  - Regulárnymi gramatikami.
  - Konečnými automatmi.
  - Regulárnymi výrazmi.
- Aké sú rôzne vlastnosti triedy regulárnych jazykov?

Veta

Nech L_1 a L_2 sú regulárne jazyky. Potom zjednotenie jazykov $L_1 \cup L_2$ je tiež regulárny jazyk.

Veta

Nech L_1 a L_2 sú regulárne jazyky. Potom zreťazenie jazykov $L_1 L_2$ je tiež regulárny jazyk.

Veta

Nech L je regulárny jazyk. Potom iterácia jazyka L^* je tiež regulárny jazyk.

Všetky tri hore uvedené vety priamo vyplývajú z definície regulárnych výrazov, resp. z konštrukcie NKA, ktorý príslušné výrazy akceptuje.

Veta

Nech L je regulárny jazyk nad abecedou T. Potom doplnok jazyka L^C je regulárny jazyk.

KA akceptujúci L^C získame, ak v **úplnom** automate akceptujúcom L vymeníme akceptujúce stavy za neakceptujúce a naopak.

Veta

Nech L_1 a L_2 sú regulárne jazyky. Potom prienik jazykov $L_1 \cap L_2$ je regulárny jazyk.

Dôkaz: Nech $M_1 = (Q_1, T_1, \delta_1, q_{01}, F_1)$ a $M_2 = (Q_2, T_2, \delta_2, q_{02}, F_2)$ sú KA také, že $L(M_1) = L_1, L(M_2) = L_2$. Potom konečný automat M, akceptujúci $L(M) = L_1 \cap L_2$ je pätica $M = (Q, T, \delta, q_0, F)$, kde:

- $Q = \{(q_1, q_2) \mid q_1 \in Q_1 \wedge q_2 \in Q_2\}, Q \subseteq Q_1 \times Q_2$
- $T = T_1 \cap T_2$
- $\delta((q_1, q_2), a) = (\delta_1(q_1, a), \delta_2(q_2, a))$
- $q_0 = (q_{01}, q_{02})$
- $F = \{(q_1, q_2) \mid q_1 \in F_1 \wedge q_2 \in F_2\}$

Konečný automat je teda kombináciou M_1 a M_2 a „sleduje“, akými cestami sa môžu uberať oba automaty a akceptuje slovo len vtedy, keď by ho akceptoval M_1 a súčasne aj M_2 (teda logicky patrí do $L_1 \cap L_2$).

Príklad: Nech $M_1 = (\{q_{01}, q_{11}\}, \{0, 1\}, \delta_1, q_{01}, \{q_{11}\})$ a $M_2 = (\{q_{02}, q_{12}, q_{22}\}, \{0, 1, 2\}, \delta_2, q_{02}, \{q_{22}\})$, kde prechodové tabuľky:

δ_1	0	1
q_{01}	q_{11}	-
q_{11}	q_{11}	q_{11}

δ_2	0	1	2
q_{02}	q_{12}	q_{22}	q_{22}
q_{12}	q_{12}	q_{22}	q_{12}
q_{22}	q_{12}	q_{22}	q_{12}

Riešenie:

δ	0	1
(q_{01}, q_{02})	(q_{11}, q_{12})	-
(q_{11}, q_{12})	(q_{11}, q_{12})	(q_{11}, q_{22})
(q_{11}, q_{22})	(q_{11}, q_{12})	(q_{11}, q_{22})

Pumpovacia lema

Veta

Nech L je (nekonečný) regulárny jazyk. Potom existuje prirodzené číslo p (tzv. pumpovacia dĺžka) také, že všetky slová $w \in L$, ktorých dĺžka $|w| \geq p$, možno vyjadriť v tvare $w = xyz$, pričom:

1. *pre každé* $i \geq 0 : xy^i z \in L$;
2. $|x| \geq 0, |z| \geq 0$
3. $|y| \geq 1$;
4. $|xy| \leq p$.

Pumpovacia lema (resp. jej **obmena**) sa využíva pri dokazovaní toho, že daný nekonečný jazyk **nie je** regulárny. Pre konečné jazyky nemá zmysel čokoľvek dokazovať, tie sú regulárne automaticky!

Dôkaz PL: Majme nekonečný regulárny jazyk L. Teda existuje konečný automat $M = (Q, T, \delta, q_0, F)$, ktorý ho rozpoznáva, $L(M) = L$. Nech p je počet stavov automatu, $p = |Q|$ a $w \in L$ slovo z jazyka, ktoré $|w| \geq p$. Keďže w je z jazyka, musí platiť:

$$(q_0, w) \vdash^* (q_f, \varepsilon), q_f \in F.$$

Ak automat má p stavov, tak celý výpočet má minimálne $p + 1$ konfigurácií, niektorý zo stavov sa musel opakovať. Označme q prvý stav, ktorý sa vo výpočte zopakuje. Navyše si rozdeľme w na 3 podreťazce $x, y, z, w = xyz$ tak, že:

$$\underbrace{(q_0, xyz) \vdash ... \vdash (q,}_{(1)} \underbrace{yz) \vdash ...(q,}_{(2)} \underbrace{z) \vdash (q_f, \varepsilon)}_{(3)}$$

Časť x prevedie výpočet zo stavu q_0 do opakujúceho stavu q (1), časť y prevedie výpočet zo stavu q späť do stavu q (2) a časť z prevedie výpočet do akceptujúceho stavu q_f. Čo je zaujímavé, tak teoreticky by sa dali vytvoriť aj nasledujúce akceptujúce výpočty:

$$\underbrace{(q_0, xz) \vdash ... \vdash (q,}_{(1)} \underbrace{z) \vdash (q_f, \varepsilon)}_{(3)}$$

$$\underbrace{(q_0, xyyz) \vdash ... \vdash (q,}_{(1)} \underbrace{yyz) \vdash ...(q,}_{(2)} \underbrace{yz) \vdash ...(q,}_{(2)} \underbrace{z) \vdash (q_f, \varepsilon)}_{(3)}$$

T.j. ak je automat v stave q a na vstupe je nespracovaná časť z, na konci výpočtu skončí v akceptujúcej konfigurácii. Podobne akceptuje aj xz, kde x dostane automat zo stavu q_0 do q. Keďže y dostane automat zo stavu q späť do stavu q, aj $y^2, y^3, ...$ dostanú automat z q do q.

DKA teda akceptuje nielen xyz, ale aj $xy^iz, i \geq 0$.

Keďže časť (2), kde dôjde k zopakovaniu stavu q obsahuje minimálne 1 krok výpočtu, tak $|y| \geq 1$.

Keďže q je prvý stav, ktorý sa vo výpočte zopakuje, tak $|xy| \leq p$, pretože k prvému zopakovaniu stavu môže prísť najneskôr po p krokoch výpočtu (lebo automat má p stavov).

Ako použiť PL pre regulárne jazyky

- Pumpovacia lema pre regulárne jazyky sa používa na dôkaz toho, že nejaký jazyk **nie je regulárny.**
- Veta má tvar implikácie, t.j. ak ukážem, že pre nekonečný jazyk platí, že sa nedá pumpovať, tak ukážem, že jazyk nie je regulárny.
- Ako teda použiť PL pre jazyk L?

1. Dôkaz sporom - je daný nekonečný jazyk a predpokladajme, že je regulárny.
2. Nech p je kladné číslo. Vezmeme slovo w z jazyka L, ktoré má aspoň p symbolov, t.j. $|w| \geq p$.
3. Uvažujeme **všetky možné** rozklady slova w na 3 časti x, y, z v tomto poradí, t.j. $w = xyz$, kde $|x| \geq 0, |y| \geq 1, |xy| \leq p$.
4. Pre každý takýto rozklad ukážeme, že **existuje** také číslo i, že $xy^iz \notin L$, t.j. ak časť y zväčším i-krát, výsledok nebude patriť do jazyka L. (t.j. nedá sa pumpovať).
5. Ak sa mi také i podarí nájsť, dochádza k sporu s predpokladom, a teda jazyk L nemôže byť regulárny.

$a^i b^i, i \in \{1, 2, 3, ...\}$

Príklad: Dokážte, že $L = \{a^i b^i \mid i \in \{1, 2, ...\}\}$ nie je regulárny jazyk.

1. Dôkaz sporom - Predpokladajme, že jazyk je regulárny. Potom nech p je číslo, o ktorom hovorí PL.
2. Uvažujme slovo $a^p b^p$. Toto slovo patrí do jazyka a zároveň $|a^p b^p| = 2p \geq p$.
3. Rozdeľme toto slovo na 3 časti x, y, z. Keďže podľa vety $|xy| \leq p$, tak xy bude tvorené len znakmi a (lebo tých je presne p), a teda $x = a^r, y = a^s, z = a^{p-r-s}, r \geq 0, s \geq 1, r + s \leq p$.
4. Ak vezmeme napr $i = 2$, tak dostaneme slovo $xy^2z = a^r a^{2s} a^{p-r-s} b^p = a^{p+s} b^p$. Keďže y má dĺžku aspoň 1, tak v slove xy^2z sa nachádza určite aspoň o 1 a viac, ako b.
5. **SPOR**, pretože xy^2z nepatrí do jazyka, a teda $a^i b^i$ nie je regulárny.

ww^R - nesprávny prístup

Príklad: Dokážte, že $L = \{ww^R \mid w \in \{a,b\}^*\}$ nie je regulárny jazyk.

1. Dôkaz sporom - Predpokladajme, že jazyk je regulárny. Potom nech p je číslo, o ktorom hovorí PL.
2. Uvažujme slovo $a^p a^p$. Toto slovo patrí do jazyka a zároveň $|a^p a^p| = 2p \geq p$.
3. Podľa PL sa takéto slovo dá rozdeliť na 3 časti x, y, z,t.j $x = a^r, y = a^s, z = a^{2p-r-s}, r \geq 0, s \geq 1, r + s \leq p$, pre všetky prípustné voľby r a s.
4. Ak chceme ukázať, že jazyk nie je regulárny, potrebujeme nájsť $i \geq 0$, že $xy^iz \notin L$ pre všetky prípustné x, y, z.
5. **To sa nám nepodarí!** Pretože pre každú voľbu i sa vždy dá nájsť rozklad x, y, z, že $xy^iz \in L$.
6. Príslušný jazyk naozaj regulárny nie je. Presvedčíme sa o tom tak, že v kroku 2. zvolíme **iné slovo**.

ww^R - správny prístup

Príklad: Dokážte, že $L = \{ww^R \mid w \in \{a,b\}^*\}$ nie je regulárny jazyk.

1. Dôkaz sporom - Predpokladajme, že jazyk je regulárny. Potom nech p je číslo, o ktorom hovorí PL.
2. Uvažujme slovo $a^p b^p b^p a^p$. Toto slovo patrí do jazyka a zároveň $|a^p b^p b^p a^p| = 4p \geq p$.
3. Podľa PL sa takéto slovo dá rozdeliť na 3 časti x, y, z,t.j $x = a^r, y = a^s, z = a^{p-r-s} b^p b^p a^p, r \geq 0, s \geq 1, r + s \leq p$.
4. Nech $i = 0$. Keďže $|y| \geq 1$, tak slovo $xy^0 z = xz = a^r a^{p-r-s} b^p b^p a^p = a^{p-s} b^p b^p a^p$. Čiže oproti slovu v kroku 2 tam chýba **minimálne jedno** a medzi prvými a-čkami. A takéto slovo $a^{p-s} b^p b^p a^p$ sa **určite nedá** rozdeliť na 2 časti, ktoré sú vzájomne opačné, t.j. nikdy nebude patriť do jazyka L!
5. Tým dostávame **SPOR s predpokladom** a teda jazyk $L = \{ww^R \mid w \in \{a,b\}^*\}$ nie je regulárny.

$b^m a^n, m \geq n; m, n \in \{0, 1, 2, ...\}$

Príklad: Dokážte, že jazyk
$L = \{b^m a^n, m \geq n; m, n \in \{0, 1, 2, ...\}\}$ nie je regulárny.

1. Nech p je nejaké kladné číslo a predpokladáme, že L je regulárny.
2. Nech $w = b^p a^p$. Určite platí $w \in L$ a taktiež $|w| \geq p$.
3. Všetky možné rozklady na slová x, y, z sú:
   $x = b^r, y = b^s, z = b^{p-r-s} a^p, r \geq 0, s \geq 1, r + s \leq p$.
4. Nech $i = 0$. Potom $xy^0 z = xz = b^{p-s} a^p$. A keďže $s \geq 1$, tak v slove $xy^0 z$ je **menej** b, než a, a teda slovo **nepatrí** do jazyka L.
5. Preto jazyk L nie je regulárny.
6. **POZOR!!!** Každá iná voľba i než 0, t.j. $i = 2, 3, 4, ...$ by nedokázala nič, pretože by nafukovala b-čka a teda by slová $xy^i z$ patrili do jazyka!

a^n, n-prvočíslo

Príklad: Dokážte, že jazyk $L = \{a^n \mid n \text{ je prvočíslo }\}$ nie je regulárny.

Idea: Treba ukázať, že pre ľubovoľné číslo p vždy existuje také slovo w z jazyka (t.j. zložené z prvočíselne-veľa a-čok, ktorých je aspoň p), pre ktorého všetky prípustné rozklady x, y, z platí, že vždy existuje také číslo i, že xy^iz nepatrí do jazyka. Táto úloha je náročnejšia než predchádzajúce, pretože je potrebné vhodne zvoliť slovo w a navyše je potrebné vhodne zvoliť exponent i, pretože pre nedbanlivo zvolené i sa väčšinou dá ukázať, že rozklad x, y, z sa pumpovať dá. Preto je potrebné zvoliť i tak, aby **nebolo pochýb o tom**, že xy^iz nepatrí do jazyka, t.j. že dĺžka $|xy^iz|$ je zložené číslo.

a^n,n-prvočíslo

1. Nech p je kladné celé číslo a predpokladajme, že L je regulárny.
2. Nech $w = a^N$, kde N je prvočíslo také, že $N > p + 1$.
3. Rozklady $w = xyz$ sú:
   $x = a^r, y = a^s, z = a^{N-r-s}, r \geq 0, s \geq 1, r + s \leq p$.
4. Nech $i = |x| + |z|$. Potom slovo xy^iz má dĺžku $|xy^iz| = |x| + i|y| + |z| = |x| + (|x| + |z|)|y| + |z| = (|x| + |z|)(1 + |y|)$, t.j. súčin 2 čísiel. Keďže oba činitele sú väčšie ako 1, tak tento súčin **nemôže byť prvočíslo**, slovo xy^iz nepatrí do jazyka L a teda jazyk **nie je regulárny**.

a^n, n-prvočíslo

Zdôvodnenie:

- Súčin $(|x|+|z|)(1+|y|)$ je určite zložené číslo vtedy, ak sú oba činitele aspoň 2. A to sú, pretože:
  - $(1+|y|)$ je určite aspoň 2, pretože $|y| \geq 1$.
  - $(|x|+|z|)$: keďže $|x| \geq 0$, musíme ukázať, že $|z| \geq 2$. A to je pravda kvôli voľbe slova $w = a^N$ tak, že $N > p+1$, lebo $N \geq p+2$ a keďže $|xy| \leq p$ a $|xyz| \geq p+2$, tak z toho $|z| \geq 2$.
- Nech by bol teda rozklad slova w na tri časti x, y, z **akýkoľvek prípustný**, vždy nájdeme takú voľbu i - podľa dĺžky segmentov x, z tak, že $i = |x|+|z|$, že dĺžka slova xy^iz je zložené číslo, a teda nepatrí do jazyka L, lebo tam patria **len také slová**, ktorých dĺžka je prvočíslo.

a^n, n-zložené číslo

- Podobne by sa dalo ukázať, že ani jazyk $L = \{a^n \mid n \text{ je zložené číslo }\}$ nie je regulárny.
- Oveľa jednoduchšie je však využiť vlastnosť regulárnych jazykov spomínanú na slajde č. 73, ktorá vraví, že ak je jazyk regulárny, tak jeho doplnok je tiež regulárny.
- Obmena tvrdenia, ktorá je rovnako pravdivá, hovorí, že ak jazyk **nie je regulárny**, tak ani jeho doplnok **nie je regulárny**.
- A keďže doplnok jazyka $L = \{a^n \mid n \text{ je zložené číslo }\}$ je práve jazyk $L = \{a^n \mid n \text{ je prvočíslo }\}$, o ktorom sme pred chvíľou dokázali, že nie je regulárny, tak potom ani náš daný jazyk $L = \{a^n \mid n \text{ je zložené číslo }\}$ nie je regulárny.

$a^{n^2}, n \in \{1, 2, 3, ...\}$

Príklad: Dokážte, že jazyk $L = \{a^{n^2}, n \in \{1, 2, 3, ...\}\}$ nie je regulárny.

1. Nech p je kladné celé číslo a L je regulárny.
2. Nech $w = a^{p^2}$.
3. Rozklady $w = xyz$ sú
   $x = a^r, y = a^s, z = a^{p^2-r-s}, r \geq 0, s \geq 1, r + s \leq p$.
4. Nech $i = 2$. Potom slovo $xy^2z = a^r a^{2s} a^{p^2-r-s} = a^{p^2+s}$.
   Aby xy^2z nepatrilo do jazyka, musí platiť, že $p^2 + s$ nie je druhá mocnina nejakého čísla.

$a^{n^2}, n \in \{1, 2, 3, ...\}$

4. Je $p^2 + s$ druhá mocnina nejakého čísla?
    - Určite $p^2 < p^2 + s$, keďže $s \geq 1$.
    - Ak p^2 je druhá mocnina nejakého čísla, tak ďalšia v poradí bude $(p+1)^2 = p^2 + 2p + 1$
    - Avšak $p^2 + s < p^2 + 2p + 1$, pretože $1 \leq s \leq p$ a $p \geq 1$.
    - To znamená, že $p^2 + s$ **nemôže byť druhá mocnina žiadneho čísla**, pretože najbližšie druhé mocniny sú p^2 a $p^2 + 2p + 1$ a všetky hodnoty $p^2 + s$ ležia **striktne medzi nimi**.
5. Preto pre $i = 2$ slovo xy^2z nepatrí do jazyka L a jazyk L teda **nie je regulárny.**

Použitá literatúra

Dedera, Ľ: Počítačové jazyky a ich spracovanie.

Linz, P.: An Introduction to Formal Languages and Automata.

Molnár, Ľ.: Gramatiky a jazyky.