Jazyky, konečné automaty

- Základní pojmy teorie formálních jazyků (abeceda, slovo, jazyk)
- Konečný automat
  - Definice
  - Použití
  - Vlastnosti rozpoznávaných jazyků

Symboly, formální abeceda

Cíl: Navrhnout matematické pojmy modelující vstupy a výstupy algoritmických výpočtů.

Pod pojmem abeceda rozumíme konečnou množinu Σ.
Prvky Σ nazýváme symboly.

Příklady abeced:

$\Sigma_1 = \{0, 1\}$ - binární abeceda

$\Sigma_2 = \{0, 1, \ldots, 9, A, \ldots, F\}$ - hexadecimální abeceda

$\Sigma_3 =$ všechny ASCII znaky

$\Sigma_4 = \{\leftarrow, \rightarrow, \uparrow, \downarrow\}$ - symboly reprezentují 4 základní směry

Formální jazyk

Slovo nad abecedou Σ je konečná posloupnost prvků ze Σ.
Slovo nulové délky nazýváme prázdné slovo, značíme jej ε.

Σ^* … množina všech slov nad abecedou Σ
$\Sigma^+ = \Sigma^* - \{\varepsilon\}$ … množina všech neprázdných slov nad Σ

Jazyk nad abecedou Σ je každá podmnožina Σ^*.

Příklad:

$\Sigma = \{a, b\}$

$L_1 = \{\varepsilon, a, b, aa, ab, ba, bb\}$ .. slova délky maximálně 2
$L_2 = \{a, ab, abb, abbb, \ldots\}$ .. slova obsahující a pouze jako jednoznakový prefix

Další příklady jazyků

- Genetické kódy
- Zdrojové soubory programů (bez syntaktických chyb)
- Gramaticky správné věty přirozeného jazyka
- Posloupnosti tahů, pomocí kterých projdeme bludiště

Slova, operace zřetězení

Pro označení slov používáme proměnné $u, v, w, \dots$
Pro symboly proměnné $a, b, c, \dots$

$|u|$ .. délka slova u

Zřetězení slov $u = a_1 a_2 \cdots a_m$, $v = b_1 b_2 \cdots b_n$

$$u \cdot v = uv = a_1 \cdots a_m b_1 \cdots b_n$$

k-násobné zřetězení:

$$u^k = \overbrace{u \cdots u}^{k}$$

$$u^0 = \varepsilon, \quad u^1 = u$$

Konečný automat

- Nejjednodušší, fundamentální model výpočetního systému.
- Konečný = konečně stavový.
- Výskyt i v jiných oblastech než jsou počítače (modelování systémů, procesů, organismů).

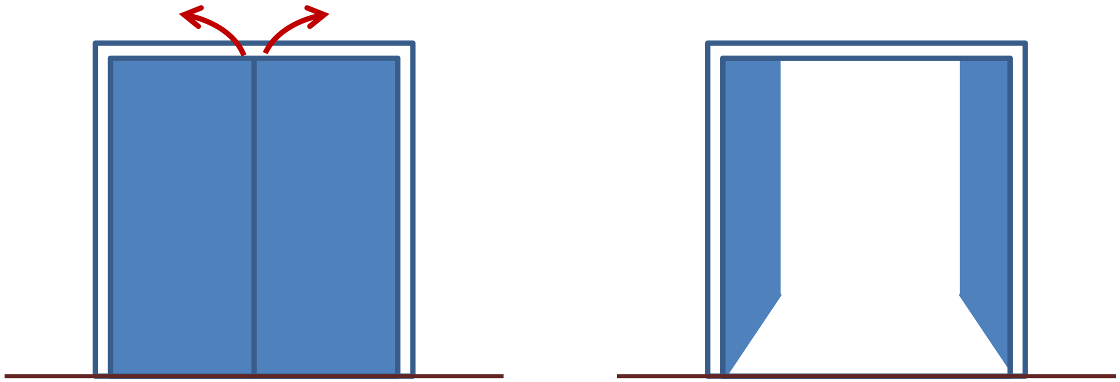
Příklady použití:

- hledání vzorků v textu
- lexikální analýza v překladačích
- regulární výrazy

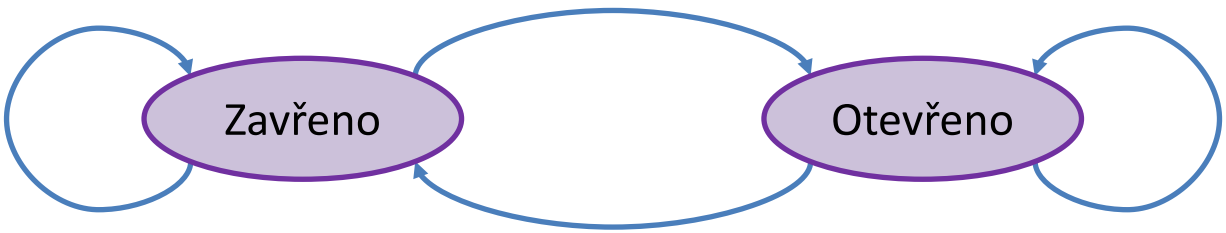
$$[\wedge a, e, i, o, u, y][a, e, i, o, u, y]+$$

Ukázka modelování systému

Vstupní automatické dveře, které se otevírají pouze dovnitř.



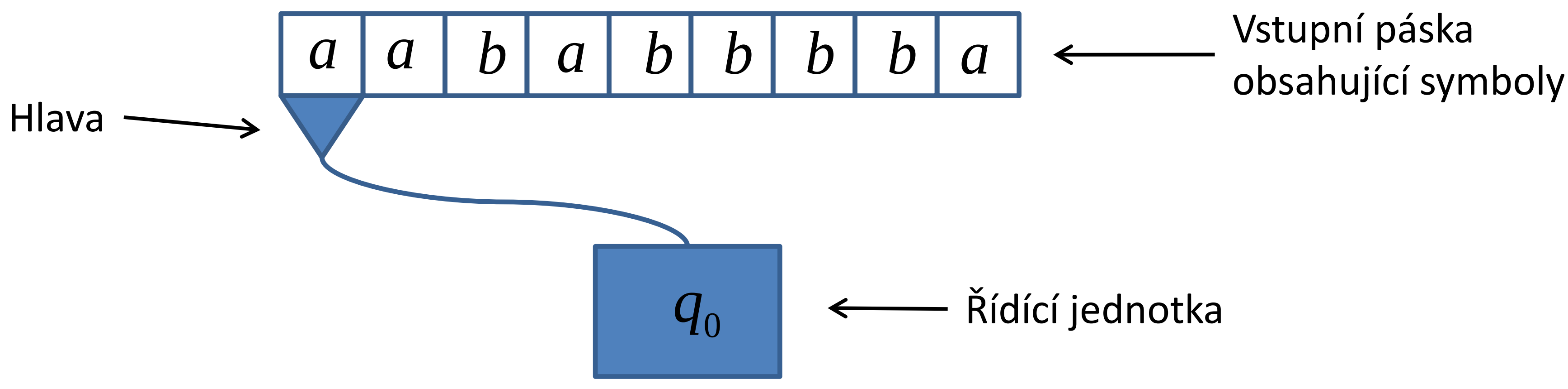
Přechody mezi stavy otevřeno/zavřeno:



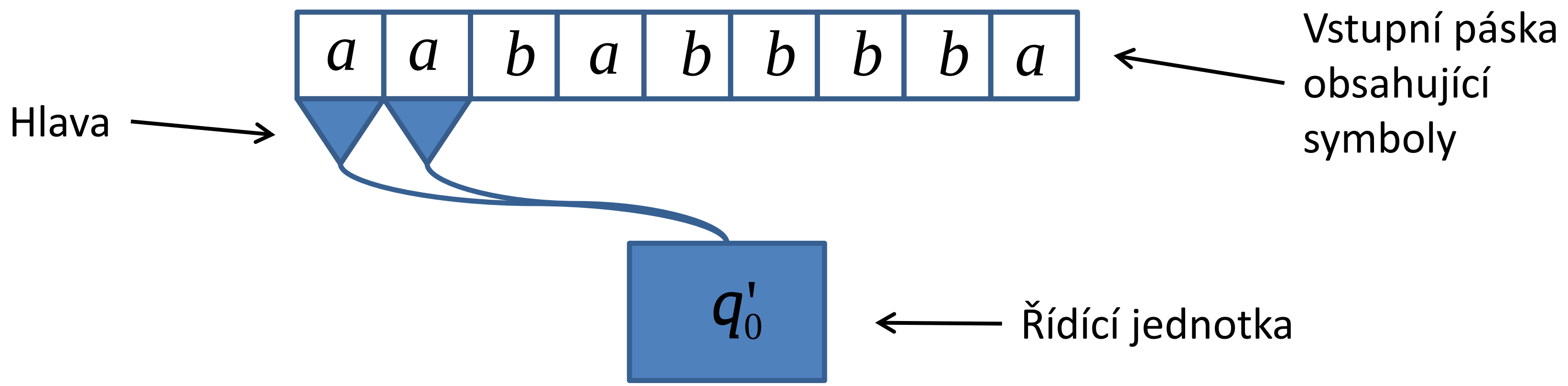
Konečný automat

Konečný automat je pětice $A = (Q, \Sigma, \delta, q_0, F)$, kde

- Q je konečná množina stavů
- Σ je abeceda
- $\delta : Q \times \Sigma \to Q$ je přechodová funkce
- $q_0 \in Q$ je počáteční (iniciální) stav
- $F \subseteq Q$ je množina přijímacích (koncových) stavů

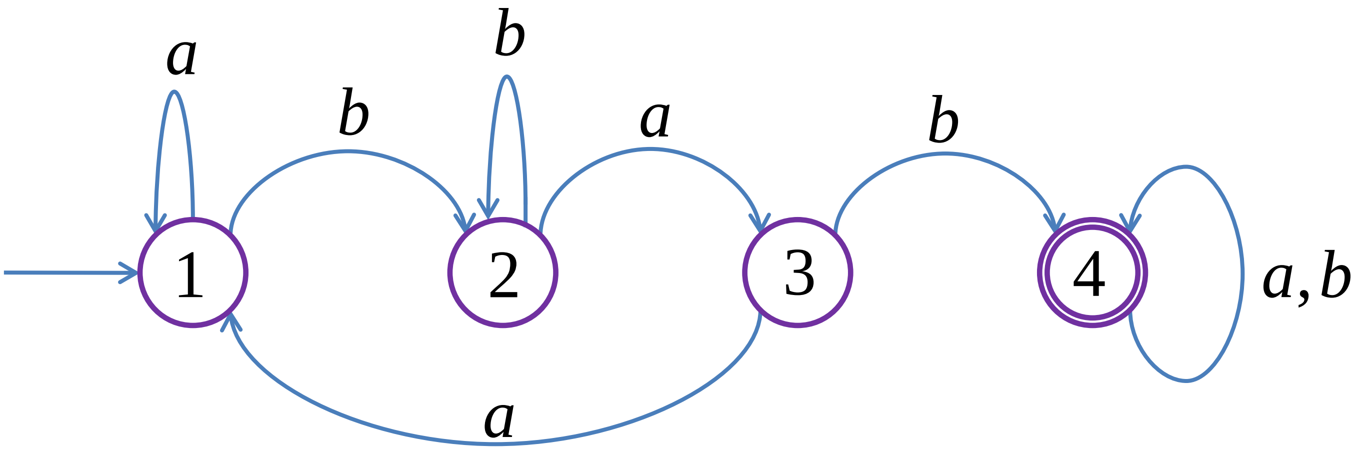


Výpočet konečného automatu



1. Řídící jednotka ve stavu q_0, hlava na levém konci pásky.
2. Opakujeme, dokud automat nepřečte celý vstup:
   - Stav se změní na $\delta(q, x)$, kde q je aktuální stav a x aktuální symbol čtený hlavou.
   - Hlava se posune o jedno pole vpravo.
3. Pokud automat skončí v přijímacím stavu, je vstupní slovo přijato, jinak je zamítnuto.

Reprezentace grafem nebo tabulkou



	a	b
→1	1	2
2	3	2
3	1	4
←4	4	4

$$Q = \{1,2,3,4\}$$
$$\Sigma = \{a,b\}$$
$$q_0 = 1$$
$$F = \{4\}$$

Konfigurace

- Konfigurace konečného automatu $A = (Q, \Sigma, \delta, q_0, F)$ je dvojice (q, w), kde $q \in Q$ a $w \in \Sigma^*$.

- $K = (q, aw) \wedge K' = (q', w) \wedge \delta(q, a) = q'$
  K' bezprostředně následuje po K

- Výpočet automatu je posloupnost konfigurací $K_0, \ldots, K_n$ pro kterou platí:

  $K_0 = (q_0, w)$

  K_{i+1} bezprostředně následuje po K_i (pro $\forall i = 0, \ldots, n-1$)

Jazyk rozpoznávaný konečným automatem

- Výpočet $K_0, \ldots, K_n$ je přijímající, pokud $K_n = (q, \varepsilon)$, kde $q \in F$.

- V takovém případě automat přijímá slovo w.

- Jazyk rozpoznávaný (přijímaný) konečným automatem A

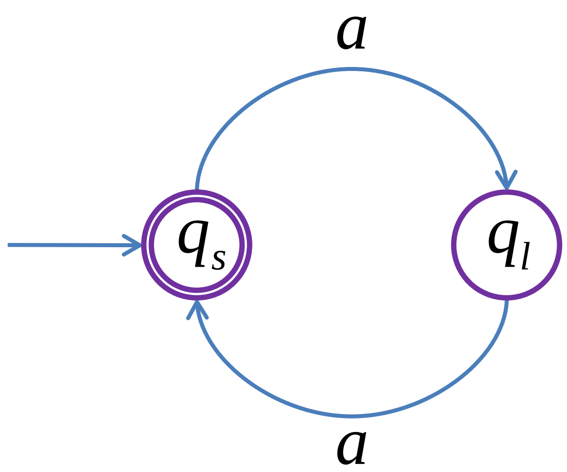
  $$L(A) = \{w \mid A \text{ přijímá } w\}$$

- Jazyk L je regulární, pokud existuje konečný automat A, který jej rozpoznává.

  $$L = L(A)$$

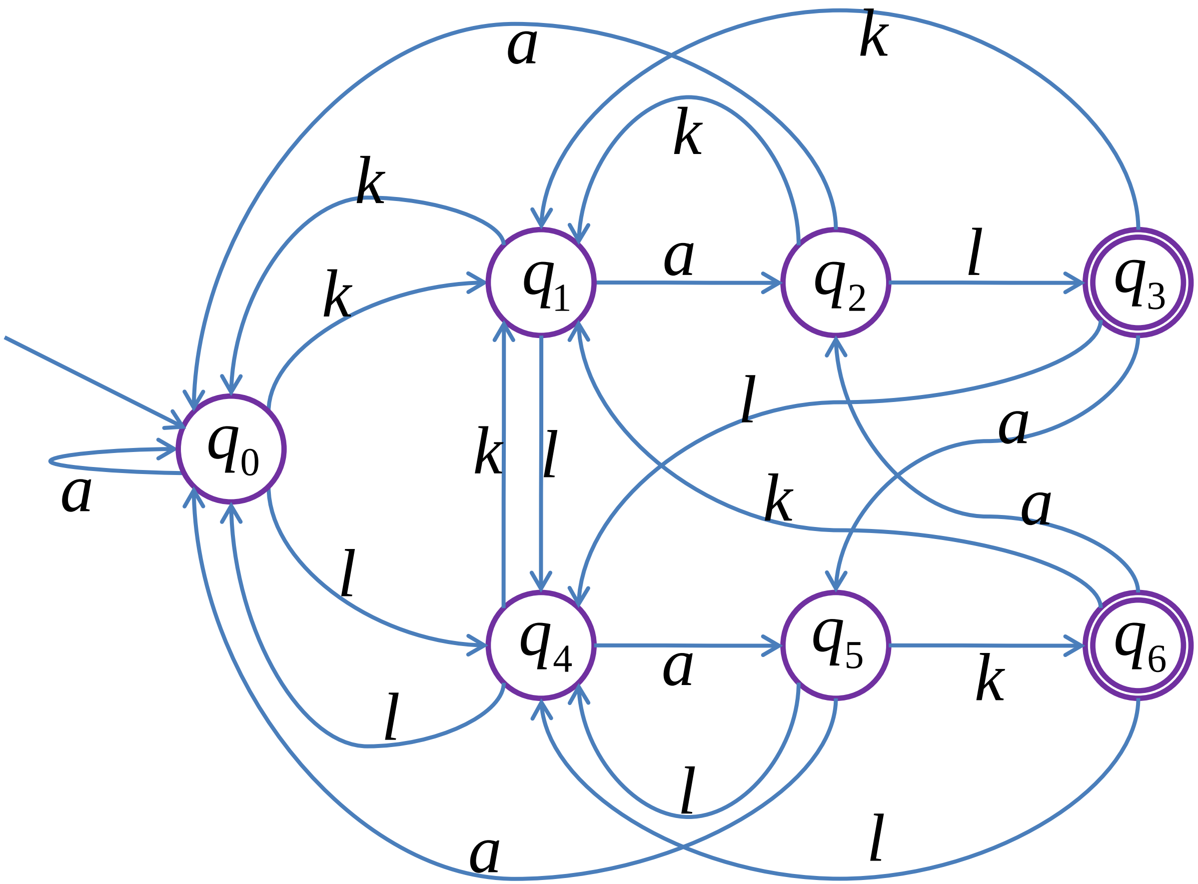
Příklad č.1

Jaký jazyk rozpoznává následující konečný automat A_1?



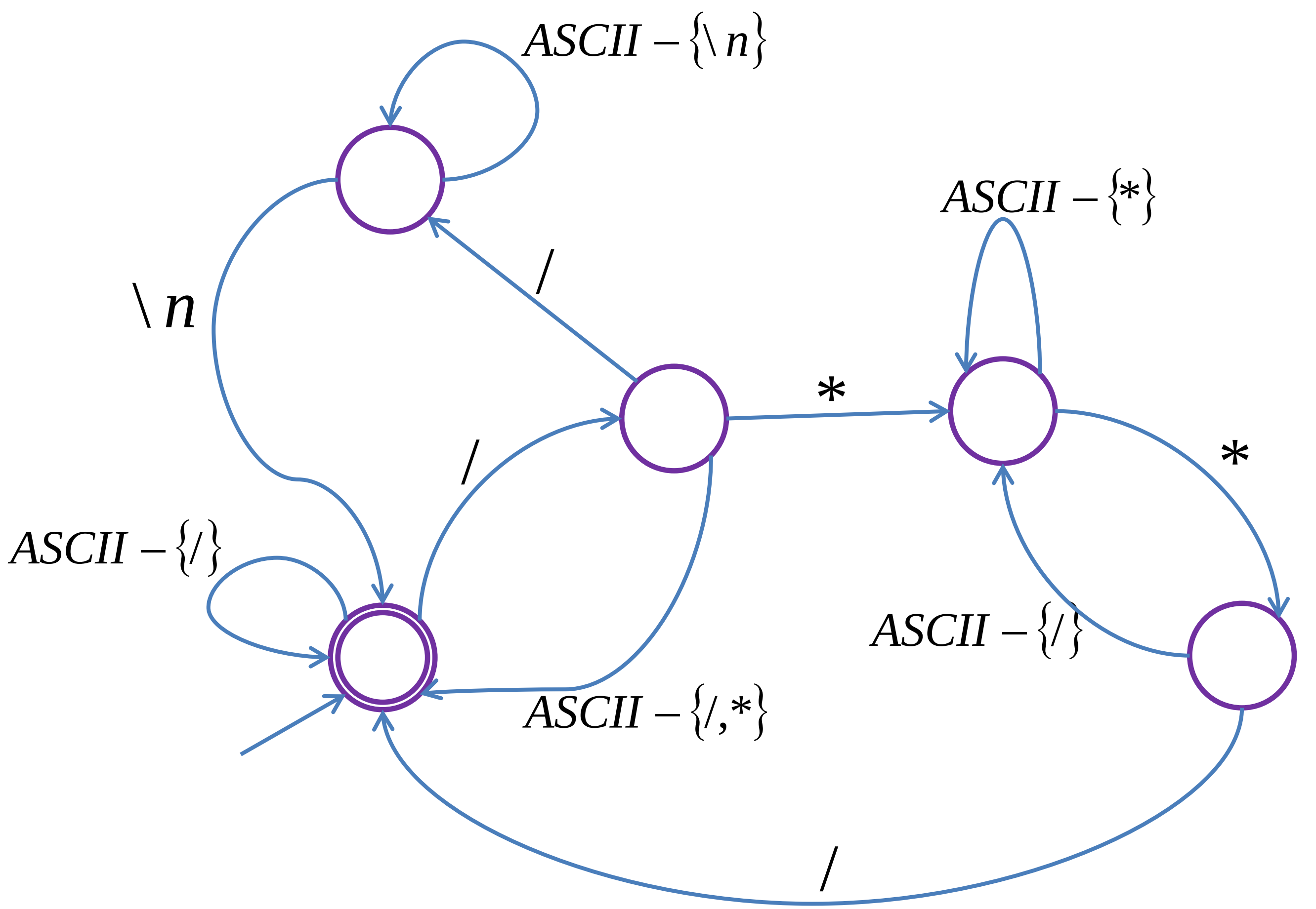
Příklad č.2

Jaký jazyk rozpoznává následující konečný automat A_2?



Příklad č.3

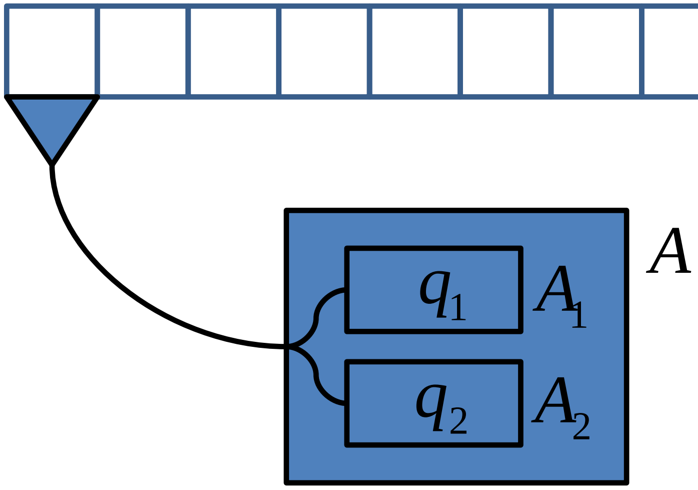
Jaký jazyk rozpoznává následující konečný automat A_3?



Skládání automatů

Věta: Nechť $L_1, L_2 \subseteq \Sigma^*$ jsou regulární jazyky. Pak je také jazyk $L_1 \cup L_2$ regulární.

$$L_1 = L(A_1) \qquad A_1 = (Q_1, \Sigma, \delta_1, q_{01}, F_1)$$
$$L_2 = L(A_2) \qquad A_2 = (Q_2, \Sigma, \delta_2, q_{02}, F_2)$$



Skládání automatů

Definujeme automat $A = (Q, \Sigma, \delta, q_0, F)$:

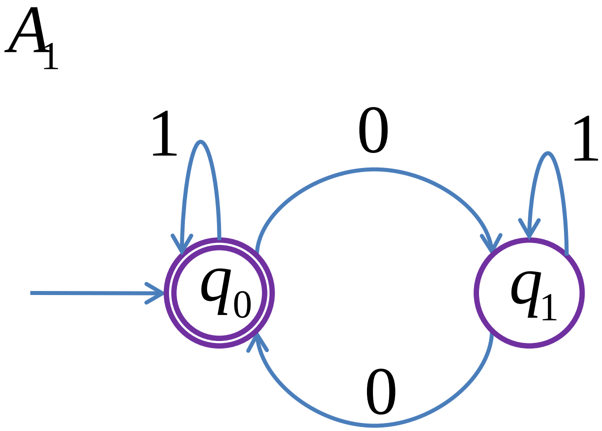
$$Q = Q_1 \times Q_2$$
$$\delta\big((q_1, q_2), a\big) = \big(\delta_1(q_1, a), \delta_2(q_2, a)\big)$$
$$q_0 = (q_{01}, q_{02})$$
$$F = \big(F_1 \times Q_2\big) \cup \big(Q_1 \times F_2\big)$$

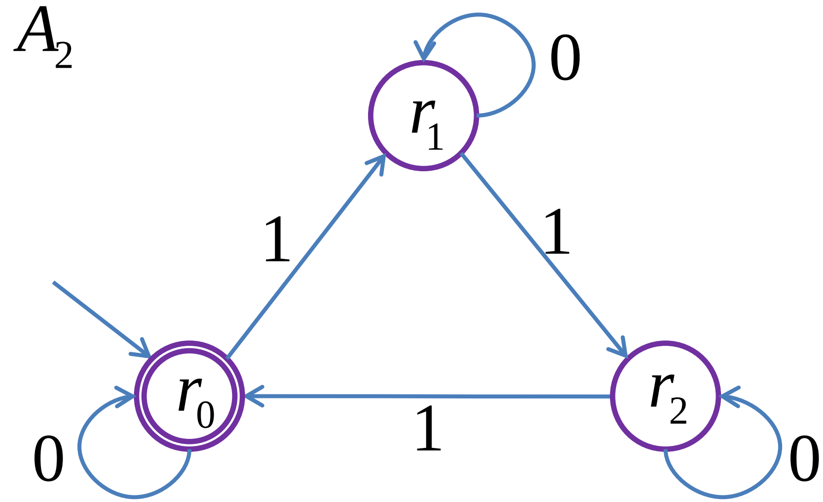
Modifikace věty pro průnik jazyků:

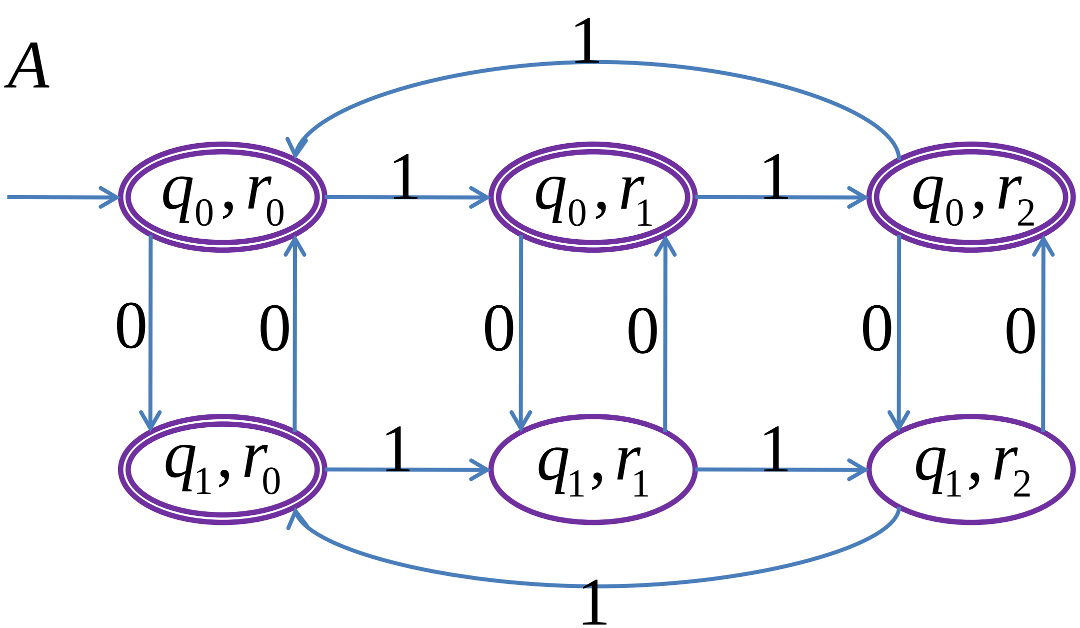
Nechť $L_1, L_2 \subseteq \Sigma^*$ jsou regulární jazyky. Pak je také jazyk $L_1 \cap L_2$ regulární.

$$F = F_1 \times F_2$$

Příklad sjednocení automatů







Další vlastnosti

Věta: Nechť L je regulární jazyk. Pak je také doplněk

$$\overline{L} = \Sigma^* - L$$ regulární.

- V původním automatu prohodíme přijímací a nepřijímací stavy.

Nechť je dán konečný automat A.

Jak poznat, zda

- Nepřijímá žádné slovo, tj. že $L(A) = \emptyset$
- Přijímá všechna slova, tj. že $L(A) = \Sigma^*$

Řešení: zkoumáme dosažitelnost přijímacích stavů.

Ekvivalence konečných automatů

Mějme dva konečné automaty A_1, A_2. Jakým způsobem můžeme zjistit, zda rozpoznávají stejný jazyk?

$$L_1 = L(A_1), \quad L_2 = L(A_2)$$

$$L_1 = L_2\ ???$$

<u>Pozorování</u>:

$$L_1 \subseteq L_2 \Leftrightarrow L_1 \cap \overline{L_2} = \emptyset$$

$$L_2 \subseteq L_1 \Leftrightarrow \overline{L_1} \cap L_2 = \emptyset$$

$$\Rightarrow \quad L_1 = L_2 \Leftrightarrow \left(L_1 \cap \overline{L_2}\right) \cup \left(\overline{L_1} \cap L_2\right) = \emptyset$$

Pumping Lemma

Věta: Nechť L je regulární jazyk. Pak existuje n takové, že každé slovo $z \in L, |z| \geq n$ lze psát ve tvaru $z = uvw$, kde $|uv| \leq n, |v| \geq 1$ a $\forall i \in \mathbf{N}$ je $uv^i w \in L$.

Použití:

Dokážeme, že jazyk $L = \{a^i b^i \mid i \in \mathbf{N}\}$ není regulární.

Pumping Lemma

Důsledek předchozího výsledku:

$$\Sigma = \{a, b\}$$
$$L = \{w \in \Sigma^* : |w|_a = |w|_b\}$$

$|w|_a$ .. počet symbolů a ve slově w

- Jazyk L není regulární.

$$L_1 = \{a^i b^j : i, j \in \mathbf{N}\} \qquad L_2 = \{a^i b^i : i \in \mathbf{N}\}$$

$$L_2 = L \cap L_1$$