# Konečné automaty

**Definice 1.** **Deteministický konečný automat** (Deterministic Finite Automaton, DFA) $\mathcal{M}$ je pětice $(Q, \Sigma, \delta, q_0, F)$, kde

- $Q$ je neprázdná konečná množina **stavů**.
- $\Sigma$ je konečná **vstupní abeceda**.
- $\delta : Q \times \Sigma \to Q$ je parciální **přechodová funkce.**
- $q_0 \in Q$ je **počáteční (iniciální) stav.**
- $F \subseteq Q$ je množina **koncových (akceptujících) stavů.**

# Příklad a zápis tabulkou

# Zápis grafem

# Výpočet konečného automatu

**Rozšířená přechodová funkce** $\hat{\delta}$: $Q \times \Sigma^* \to Q$ je parciální funkce definovaná induktivně vzhledem k délce slova ze $\Sigma^*$:

- $\hat{\delta}(q, \varepsilon) = q$ pro každý stav $q \in Q$.
- $\hat{\delta}(q, wa) = \begin{cases} \delta(\hat{\delta}(q, w), a) & \text{je-li } \hat{\delta}(q, w) \text{ i } \delta(\hat{\delta}(q, w), a) \text{ definováno,} \\ \perp & \text{jinak.} \end{cases}$

Slovo $w$ je **akceptováno** automatem $\mathcal{M}$ právě když $\hat{\delta}(q_0, w) \in F$.

Slovo $w$ je **zamítáno** automatem $\mathcal{M}$ právě když $\hat{\delta}(q_0, w) \notin F$.

Jazyk **přijímaný** (**akceptovaný**, **rozpoznávaný**) automatem $\mathcal{M}$ je

$$L(\mathcal{M}) = \{w \in \Sigma^* \mid \hat{\delta}(q_0, w) \in F\}.$$

Jazyk, který je rozpoznatelný (nějakým) deterministickým konečným automatem, nazveme **regulární**.

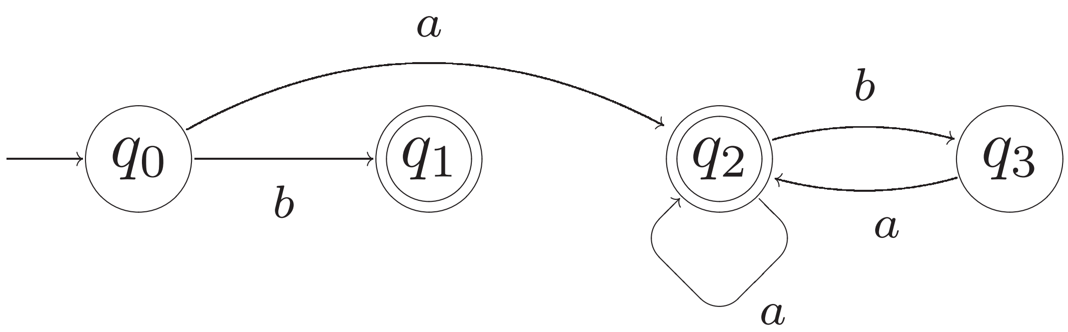
**Ekvivalenci** deterministických konečných automatů definujeme podobně jako v případě gramatik: automaty $\mathcal{M}$ a $\mathcal{M}'$ jsou ekvivalentní, pokud $L(\mathcal{M}) = L(\mathcal{M}')$.

# Parcialita přechodové funkce

Přechodová funkce $\delta$ zavedena jako parciální.
Parcialita přechodové funkce nemá podstatný vliv na výpočetní sílu konečných automatů.

**Lemma 1.** Ke každému DFA $\mathcal{M}$ existuje ekvivalentní DFA $\mathcal{M}'$ s totální přechodovou funkcí.

Idea důkazu



**Důkaz.** Nechť $\mathcal{M} = (Q, \Sigma, \delta, q_0, F)$.
Nechť automat $\mathcal{M}'$ je definován předpisem
$\mathcal{M}' = (Q \cup \{p\}, \Sigma, \delta', q_0, F)$, kde $p \notin Q$ a

$$\delta'(q,a) = \begin{cases} \delta(q,a) & \text{je-li } \delta(q,a) \text{ definováno,} \\ p & \text{jinak.} \end{cases}$$

Zejména $\delta'(p,a) = p$ pro každé $a \in \Sigma$.

*Důkaz korektnosti:*

- $\mathcal{M}$' má totální přechodovou funkci – zřejmé z definice $\mathcal{M}'$.
- $\mathcal{M}$ a $\mathcal{M}$' jsou ekvivalentní – dokážeme.

Indukcí k délce slova ověříme, že pro každé $q \in Q$ a $w \in \Sigma^*$ platí

$$\widehat{\delta'}(q,w) = \begin{cases} \hat{\delta}(q,w) & \text{je-li } \hat{\delta}(q,w) \text{ definováno,} \\ p & \text{jinak.} \end{cases}$$

Jelikož $p \notin F$, platí $L(\mathcal{M}) = L(\mathcal{M}')$. $\square$

# Konstrukce konečných automatů

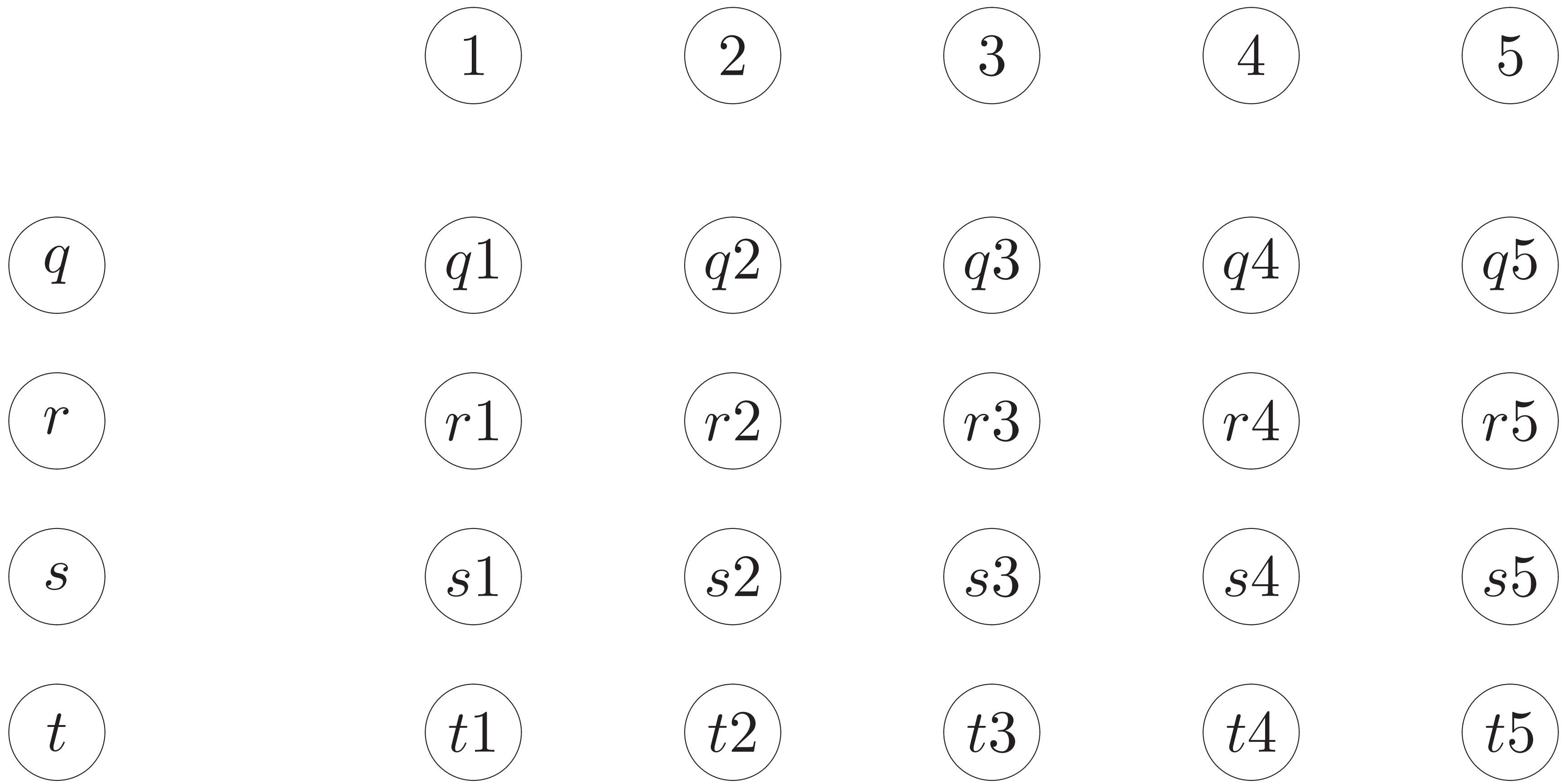
Máme za úkol sestrojit automat rozpoznávající jazyk

$$L = \{w \in \{a,b\}^* \mid w \text{ obsahuje podslovo } abaa\}$$

Označení stavů automatu zvolíme tak, aby bylo patrné, jaká část požadovaného podslova $abaa$ již byla automatem přečtena:

# Příklad

$\{w \in \{a,b\}^* \mid w$ obsahuje podslovo $abaa \ \wedge\ (w = b \vee w$ začíná i končí na $a$ a mezi dvěma výskyty $b$ je alespoň jedno $a)\}$



# Synchronní paralelní kompozice automatů

Pro dané automaty $\mathcal{M}_1$ a $\mathcal{M}_2$ umožňuje sestrojit automat rozpoznávající **průnik** (**sjednocení**, **rozdíl**) jazyků $L(\mathcal{M}_1)$ a $L(\mathcal{M}_2)$.

Nechť $\mathcal{M}_1 = (Q_1, \Sigma, \delta_1, q_1, F_1)$, $\mathcal{M}_2 = (Q_2, \Sigma, \delta_2, q_2, F_2)$
a **přechodové funkce** $\delta_1, \delta_2$ **jsou totální**.

Definujeme DFA $\mathcal{M}_3 = (Q_3, \Sigma, \delta_3, q_3, F_3)$, kde

- $Q_3 = Q_1 \times Q_2 = \{(p, q) \mid p \in Q_1, q \in Q_2\}$
- $F_3 = F_1 \times F_2 = \{(p, q) \mid p \in F_1, q \in F_2\}$
- $q_3 = (q_1, q_2)$
- $\delta_3((p, q), a) = (\delta_1(p, a), \delta_2(q, a))$

**Tvrzení:** $L(\mathcal{M}_3) = L(\mathcal{M}_1) \cap L(\mathcal{M}_2)$

**Důkaz.** Nejprve dokážeme toto tvrzení:

$$\widehat{\delta_3}((q_1, q_2), w) = (p, q) \iff \widehat{\delta_1}(q_1, w) = p \land \widehat{\delta_2}(q_2, w) = q$$

Důkaz se provede indukcí vzhledem k $|w|$.

- **Základní krok $|w| = 0$:**
  Z definice $\widehat{\delta_3}((q_1, q_2), \varepsilon) = (q_1, q_2)$, $\widehat{\delta_1}(q_1, \varepsilon) = q_1$, $\widehat{\delta_2}(q_2, \varepsilon) = q_2$.
  Pro $w = \varepsilon$ je tedy ekvivalence platná.

- **Indukční krok:** Nechť $w = va$, kde $v \in \Sigma^*$, $a \in \Sigma$. Platí:
  $\widehat{\delta_3}((q_1, q_2), va) = (p, q) \iff$

  $\widehat{\delta_3}((q_1, q_2), v) = (r, s) \ \land \ \delta_3((r, s), a) = (p, q) \iff$

  $\widehat{\delta_1}(q_1, v) = r \ \land \ \widehat{\delta_2}(q_2, v) = s \land \delta_1(r, a) = p \ \land \ \delta_2(s, a) = q \iff$

  $\widehat{\delta_1}(q_1, va) = p \ \land \ \widehat{\delta_2}(q_2, va) = q$

Nyní již lze snadno dokázat vlastní tvrzení věty:

$w \in L(\mathcal{M}_3) \iff$

$\widehat{\delta_3}((q_1, q_2), w) = (p, q)$ kde $p \in F_1$ a $q \in F_2 \iff$

$\widehat{\delta_1}(q_1, w) = p \wedge \widehat{\delta_2}(q_2, w) = q \iff$

$w \in L(\mathcal{M}_1) \cap L(\mathcal{M}_2)$. $\square$

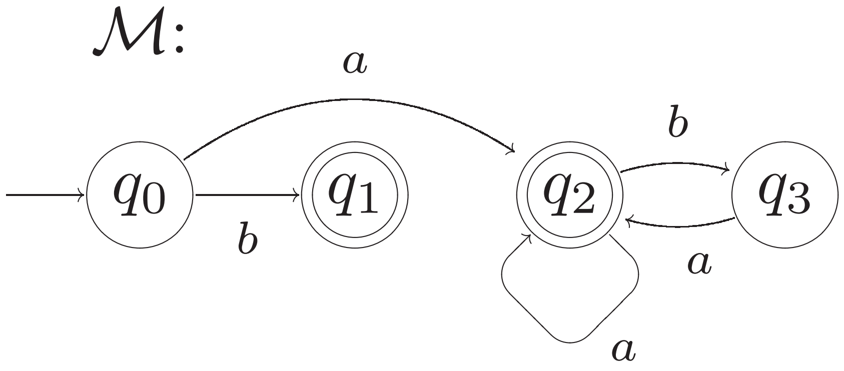
Modifikace pro **sjednocení**, tj. $L(\mathcal{M}_3) = L(\mathcal{M}_1) \cup L(\mathcal{M}_2)$:

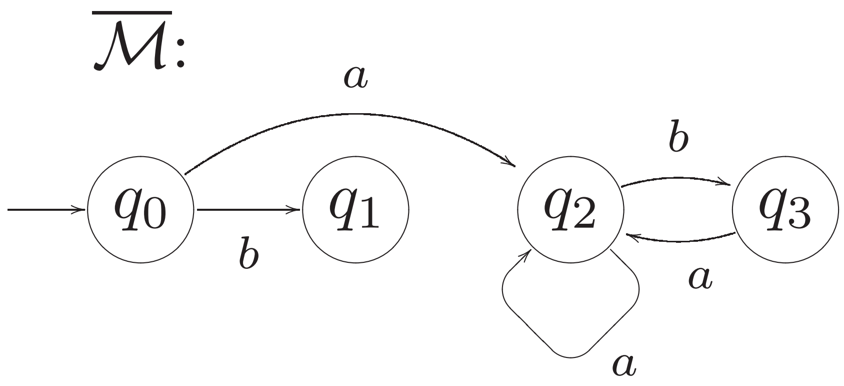
**DÚ:** Modifikujte konstrukci tak, aby platilo $L(\mathcal{M}_3) = L(\mathcal{M}_1) \smallsetminus L(\mathcal{M}_2)$.

# Automat pro komplement

K automatu $\mathcal{M} = (Q, \Sigma, \delta, q_0, F)$ s **totální přechodovou funkcí** sestrojíme automat $\overline{\mathcal{M}}$ rozpoznávající jazyk co–$L(\mathcal{M})$ jako

$$\overline{\mathcal{M}} = (Q, \Sigma, \delta, q_0, Q \smallsetminus F).$$



# Limity konečných automatů

$$L = \{a^n b^n \mid n \geq 0\} = \{\varepsilon, ab, aabb, aaabbb, aaaabbbb \ldots\}$$

$$a \ a \ a \ a \ a \ b \ b \ b \ b \ b$$

Předpokládejme, že existuje automat $\mathcal{M}$ příjímající jazyk $L$.
Nechť $\mathcal{M}$ má $k$ stavů.

Uvažme výpočet $\mathcal{M}$ na slově $a^n b^n$ kde $n > k$.

$$aaaaaaaaaaaaaaaaaaa\ bbbbbbbbbbbbbbbbbbb$$

Protože $n > k$, musí existovat (z Dirichletova principu) stav $p$ takový, že při čtení iniciální posloupnosti symbolů $a$ projde automat stavem $p$ (alespoň) dvakrát.

$$aaaaaaaa\ aaaaaaa\ aaaabbbbbbbbbbbbbbbbbbb$$

Platí

$$\hat{\delta}(q_0, x) = p \qquad \hat{\delta}(p, y) = p \qquad \hat{\delta}(p, z) = r \in F$$

Pak ale

$$\hat{\delta}(q_0, xz) \;=\; \hat{\delta}(\hat{\delta}(q_0, x), z) \;=\; \hat{\delta}(p, z) \;=\; r \in F$$

$$aaaaaaaa\ \ aaaabbbbbbbbbbbbbbbbbbb$$

## Analogicky můžeme "vsunout" slovo $y$

$$
\begin{aligned}
\hat{\delta}(q_0, xyyyz) &= \hat{\delta}(\hat{\delta}(\hat{\delta}(\hat{\delta}(\hat{\delta}(q_0, x), y), y), y), z) \\
&= \hat{\delta}(\hat{\delta}(\hat{\delta}(\hat{\delta}(p, y), y), y), z) \\
&= \hat{\delta}(\hat{\delta}(\hat{\delta}(p, y), y), z) \\
&= \hat{\delta}(\hat{\delta}(p, y), z) \\
&= \hat{\delta}(p, z) \\
&= r \in F
\end{aligned}
$$

**Lemma 2. [o vkládání, pumping lemma]** Nechť $L$ je regulární jazyk. Pak existuje $n \in \mathbb{N}$ takové,

že libovolné slovo $w \in L$ délky alespoň $n$ lze psát ve tvaru
$w = xyz$, kde $|xy| \leq n$, $y \neq \varepsilon$ a $xy^iz \in L$ pro každé $i \in \mathbb{N}_0$.

**Důkaz.** Nechť DFA $\mathcal{M} = (Q, \Sigma, \delta, q_0, F)$ rozpoznává jazyk $L$.
Položme $n = card(Q)$.
Pro libovolné slovo $w \in L$ délky alespoň $n$ platí, že automat $\mathcal{M}$ projde při akceptování slova $w$ (alespoň) dvakrát stejným stavem.

Slovo $w$ se tedy můžeme rozdělit na tři části: $w = xyz$, kde $y \neq \varepsilon$ a $\hat{\delta}(q_0, x) = p$, $\hat{\delta}(p, y) = p$ a $\hat{\delta}(p, z) = r \in F$. Je zřejmé, že ke zopakování nějakého stavu dojde nejpozději po zpracování prvních $n$ znaků a tedy dostáváme $|xy| \leq n$.

Dále $\hat{\delta}(p, y^i) = p$ pro libovolné $i \in \mathbb{N}_0$, proto také $\hat{\delta}(q_0, xy^iz) = r$, tj. $xy^iz \in L(\mathcal{M})$ pro každé $i \in \mathbb{N}_0$. $\square$

$$L \text{ je regulární } \implies \exists n \in \mathbb{N} \;.\; \forall w \in L \;.\; \Big( |w| \geq n \implies \exists\, x, y, z \;.\; ( w = xyz \;\wedge\; y \neq \varepsilon \;\wedge\; |xy| \leq n \;\wedge\; \forall i \geq 0 \;.\; xy^i z \in L \,) \Big)$$

**Pomocí Lemmatu lze dokázat, že nějaký jazyk není regulární**

Nechť pro jazyk $L$ platí:

- pro *libovolné* $n \in \mathbb{N}$
- *existuje* takové slovo $w \in L$ délky alespoň $n$, pro které platí, že
- při *libovolném* rozdělení slova $w$ na takové tři části $x, y, z$, že $|xy| \leq n$ a $y \neq \varepsilon$
- *existuje alespoň jedno* $i \in \mathbb{N}_0$ takové, že $xy^i z \notin L$.

Pak z Lemma o vkládání plyne, že $L$ není regulární.

# Příklad důkazu ne-regularity pomocí Lemmatu o vkládání

$L = \{uc^m u^R \mid u \in \{a,b\}^*, m > 0\}$

# Myhill-Nerodova věta
# Motivace I

$L = \{w \in \{a, b\}^* \mid \#_a(w) \geq 3\}$

# Pravá kongruence

**Definice 2.20.** Nechť $\Sigma$ je abeceda a $\sim$ je ekvivalence na $\Sigma^*$.

Ekvivalence $\sim$ je **pravá kongruence (zprava invariantní)**, pokud pro každé $u, v, w \in \Sigma^*$ platí

$$u \sim v \Longrightarrow uw \sim vw$$

**Index** ekvivalence $\sim$ je počet tříd rozkladu $\Sigma^*/_\sim$.
Je-li těchto tříd nekonečně mnoho, klademe index $\sim$ roven $\infty$.

---

**Tvrzení 2.21.** Ekvivalence $\sim$ na $\Sigma^*$ je pravá kongruence právě když pro každé $u, v \in \Sigma^*$, $a \in \Sigma$ platí $u \sim v \Longrightarrow ua \sim va$.
*(Implikace $\Longrightarrow$ je triviální, implikace $\Longleftarrow$ se snadno ukáže indukcí k délce zprava přiřetězeného slova $w$.)*

# Příklad

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

$\sim$: $u \sim v \iff u$ a $v$ začínají stejným symbolem
$\sim$ má . . . třídy ekvivalence

$\sim$: $u \sim v \iff \#_a(u) = \#_a(v)$
$\sim$ má . . . tříd ekvivalence

$\sim$: $u \sim v \iff u$ a $v$ mají stejné předposlední písmeno

# Myhill-Nerodova věta
# Motivace II

# Prefixová ekvivalence

**Definice 2.25.** Nechť $L$ je libovolný (ne nutně regulární) jazyk nad abecedou $\Sigma$. Na množině $\Sigma^*$ definujeme relaci $\sim_L$ zvanou **prefixová ekvivalence pro** $L$ takto:

$$u \sim_L v \overset{def}{\iff} \forall w \in \Sigma^* : uw \in L \iff vw \in L$$

# Příklad

$L = \{w \in \{a,b\}^* \mid \#_a(w) \geq 3\}$

$\sim_L$ má . . . třídy ekvivalence

$L = \{a^n b^n \mid n \geq 0\}$

$\sim_L$ má . . . tříd ekvivalence

# Myhill-Nerodova věta

**Věta 2.28.** Nechť $L$ je jazyk nad $\Sigma$.
Pak tato tvrzení jsou ekvivalentní:

**1.** $L$ je rozpoznatelný deterministickým konečným automatem.

**2.** $L$ je sjednocením některých tříd rozkladu určeného pravou kongruencí na $\Sigma^*$ s konečným indexem.

**3.** Relace $\sim_L$ má konečný index.

**Důkaz.**

$\mathbf{1} \Longrightarrow \mathbf{2}$

$\mathbf{2} \Longrightarrow \mathbf{3}$

$\mathbf{3} \Longrightarrow \mathbf{1}$

$\square$

# $1 \Longrightarrow 2$

**Jestliže** $L$ je rozpoznatelný deterministickým konečným automatem **pak** $L$ je sjednocením některých tříd rozkladu určeného pravou kongruencí na $\Sigma^*$ s konečným indexem.

- pro daný $L$ rozpoznávány automatem $\mathcal{M}$ zkonstruujeme relaci požadovaných vlastností
- $\mathcal{M} = (Q, \Sigma, \delta, q_0, F)$, $\delta$ je totální
- na $\Sigma^*$ definujme binární relaci $\sim$ předpisem

$$u \sim v \stackrel{def}{\Longleftrightarrow} \hat{\delta}(q_0, u) = \hat{\delta}(q_0, v)$$

- ukážeme, že $\sim$ má požadované vlastnosti

$$u \sim v \overset{def}{\iff} \hat{\delta}(q_0, u) = \hat{\delta}(q_0, v)$$

– $\sim$ je ekvivalence (je reflexivní, symetrická, tranzitivní)

– $\sim$ má konečný index
*třídy rozkladu odpovídají stavům automatu*

– $\sim$ je pravá kongruence:
Nechť $u \sim v$ a $a \in \Sigma$. Pak $\hat{\delta}(q_0, ua) = \delta(\hat{\delta}(q_0, u), a) = \delta(\hat{\delta}(q_0, v), a) = \hat{\delta}(q_0, va)$ a tedy $ua \sim va$.

– $L$ je sjednocením těch tříd rozkladu určeného relací $\sim$, které odpovídají koncovým stavům automatu $\mathcal{M}$ ■

# $2 \Longrightarrow 3$

**Nechť** $L$ je sjednocením některých tříd rozkladu určeného pravou kongruencí $R$ na $\Sigma^*$ s konečným indexem.
**Pak** prefixová ekvivalence $\sim_L$ má konečný index.

- $uRv \Longrightarrow u \sim_L v$ pro všechna $u, v \in \Sigma^*$ (tj. $R \subseteq \sim_L$)

- každá třída ekvivalence relace $R$ je **celá** obsažena v nějaké třídě ekvivalence $\sim_L$
- index ekvivalence $\sim_L$ je menší nebo roven indexu ekvivalence $R$
- $R$ má konečný index $\Longrightarrow \sim_L$ má konečný index $\blacksquare$

# $3 \Longrightarrow 1$

**Nechť** prefixová ekvivalence $\sim_L$ má konečný index.
**Pak** jazyk $L$ je rozpoznatelný deterministickým konečným automatem.

Zkonstruujeme automat $\mathcal{M} = (Q, \Sigma, \delta, q_0, F)$ přijímající $L$:

- $Q = \Sigma^*/\sim_L$
  *Stavy jsou třídy rozkladu $\Sigma^*$ určeného ekvivalencí $\sim_L$.* (Konečnost!)

- $F = \{[v] \mid v \in L\}$

- $q_0 = [\varepsilon]$

- $\delta$ je definována pomocí reprezentantů: $\delta([u], a) = [ua]$
  *Definice $\delta$ je korektní, protože nezávisí na volbě reprezentanta.*

## Důkaz korektnosti, tj. $L = L(\mathcal{M})$

- $\hat{\delta}([\varepsilon], v) = [v]$ pro každé $v \in \Sigma^*$ (indukcí k délce slova $v$)

- $v \in L(\mathcal{M}) \iff \hat{\delta}([\varepsilon], v) \in F \iff [v] \in F \iff v \in L$ $\blacksquare$

# Použití Myhill-Nerodovy věty k důkazu neregularity

$$L = \{a^n b^n \mid n \geq 0\}$$

Nechť $i \neq j$. Pak $a^i \not\sim_L a^j$, protože $a^i b^i \in L$ ale $a^j b^i \notin L$.

Žádné ze slov $a^1, a^2, a^3, a^4, a^5, a^6, \ldots$ nepadnou do stejné třídy ekvivalence relace $\sim_L$.

$\sim_L$ **nemá** konečný index $\implies$
$L$ **není regulární** $(\neg 3 \implies \neg 1)$

# Použití Myhill-Nerodovy věty k důkazu regularity

$$L = \{w \in \{a,b\}^* \mid \#_a(w) \geq 3\}$$

Třídy ekvivalence relace $\sim_L$:

$$\begin{aligned} T_1 &= \{w \in \{a,b\}^* \mid \#_a(w) = 0\} \\ T_2 &= \{w \in \{a,b\}^* \mid \#_a(w) = 1\} \\ T_3 &= \{w \in \{a,b\}^* \mid \#_a(w) = 2\} \\ T_4 &= \{w \in \{a,b\}^* \mid \#_a(w) \geq 3\} \end{aligned}$$

$\sim_L$ **má** konečný index $\implies$
$L$ **je** rozpoznatelný deterministickým konečným automatem,
tj. **regulární** $(3 \implies 1)$

# Další použití Myhill-Nerodovy věty

**Věta 2.29. a 2.31.** Minimální deterministický konečný automat s totální přechodovou funkcí akceptující jazyk $L$ je určen jednoznačně až na isomorfismus (tj. přejmenování stavů). Počet stavů tohoto automatu je roven indexu prefixové ekvivalence $\sim_L$.