Formální jazyky a automaty

Deterministické konečné automaty

Jan Křetínský

Fakulta informatiky, MU Brno

Jaro 2024

Automaty

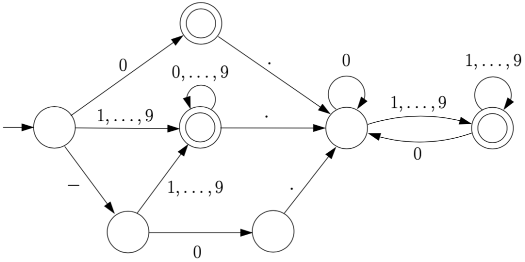
- Výpočetní zařízení

- Datová struktura

Automaty

- Výpočetní zařízení

- Datová struktura



Konečné automaty

Definice.

Deterministický konečný automat (Deterministic Finite Automaton, DFA) $\mathcal{M}$ je pětice $(Q, \Sigma, \delta, q_0, F)$, kde

- Q je neprázdná konečná množina stavů,
- Σ 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.

Výpočet konečného automatu – 1. možnost

Konfigurace je dvojice $(q, w) \in Q \times \Sigma^*$.

Krok výpočtu $\vdash$ je relace na konfiguracích:

$$(q, aw) \ \vdash \ (p, w) \quad \Longleftrightarrow \quad \delta(q, a) = p$$

$\vdash^*$ je reflexivní a tranzitivní uzávěr

Slovo w je akceptováno pokud $(q_0, w) \vdash^* (q, \varepsilon)$ kde $q \in F$

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

$$L(\mathcal{M}) = \{w \in \Sigma^* \mid (q_0, w) \vdash^* (q, \varepsilon), q \in F\}$$

Výpočet konečného automatu – 2. možnost

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

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

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

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

Ekvivalence obou možností:

- indukcí vzhledem k délce slova w

$$\hat{\delta}(q,w) = p \iff (q,w) \vdash^* (p,\varepsilon)\,,$$

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

Automaty $\mathcal{M}$ a $\mathcal{M}'$ nazveme ekvivalentní, právě když $L(\mathcal{M}) = L(\mathcal{M}')$.

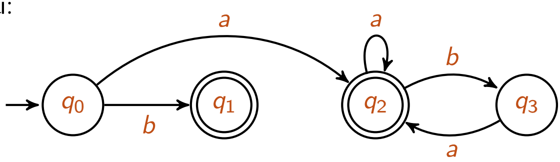
Parcialita přechodové funkce

Přechodová funkce δ byla zavedena jako parciální. Parcialita přechodové funkce nemá podstatný vliv na výpočetní sílu konečných automatů.

Lemma

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

Idea důkazu:



Důkaz. Nechť $\mathcal{M} = (Q, \Sigma, \delta, q_0, F)$. Pak $\mathcal{M}'$ definujeme 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. Nechť $\mathcal{M} = (Q, \Sigma, \delta, q_0, F)$. Pak $\mathcal{M}'$ definujeme 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:

Důkaz. Nechť $\mathcal{M} = (Q, \Sigma, \delta, q_0, F)$. Pak $\mathcal{M}'$ definujeme 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:

  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

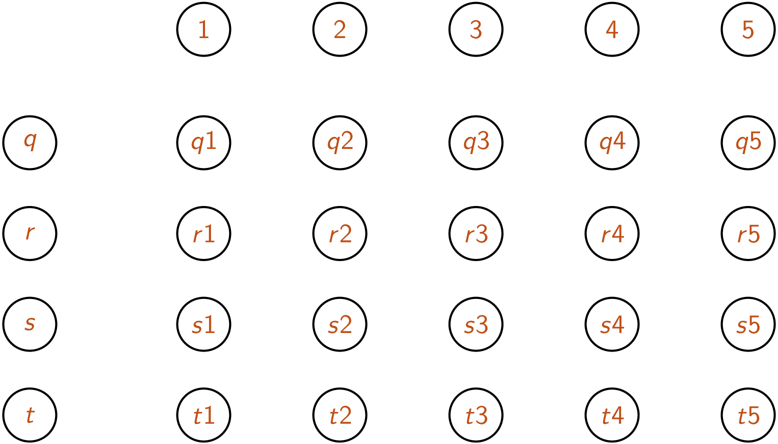
Máme za úkol sestrojit automat rozpoznávající jazyk

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

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 $abaa$ a jeho délka je dělitelná $4\}$



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** δ_1, δ_2 **jsou totální**.

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

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

Věta

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

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:

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

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

$\widehat{\delta_1}(q_1, w) = p \in F_1 \wedge \widehat{\delta_2}(q_2, w) = q \in F_2 \iff$

$w \in L(\mathcal{M}_1) \wedge w \in L(\mathcal{M}_2) \iff$

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

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

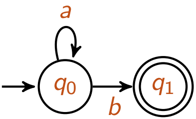
DÚ: Modifikujte konstrukci tak, aby platilo $L(\mathcal{M}_3) = L(\mathcal{M}_1) \setminus L(\mathcal{M}_2)$.

Automat pro komplement

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

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

$\mathcal{M}$:



$\overline{\mathcal{M}}$: