Vlastnosti regulárních jazyků

❖ Podobně jako u dalších tříd jazyků budeme nyní zkoumat následující vlastnosti regulárních jazyků:

- vlastnosti strukturální,
- vlastnosti uzávěrové a
- rozhodnutelné problémy pro regulární jazyky.

Strukturální vlastnosti regulárních jazyků

Konečné jazyky

Věta 3.1 Každý konečný jazyk je regulární.

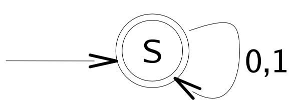
Důkaz. Nechť $L = \{w_1, w_2, \ldots, w_n\},\ w_i \in \Sigma$.

Pak $L = L(G)$, kde $G = (\{S\}, \Sigma, \{S \rightarrow w_1, S \rightarrow w_2, \ldots, S \rightarrow w_n\}, S)$. G je zřejmě gramatika typu 3.

$\square$

Opak věty 3.1 zjevně neplatí:

Příklad 3.1 Sestrojte gramatiku typu 3 generující jazyk $\{0, 1\}^*$.

Řešení: 

$\Rightarrow \quad G = (\{S\},\ \{0, 1\},\ \{S \rightarrow \varepsilon, S \rightarrow 0S, S \rightarrow 1S\},\ S)$

Pumping lemma

Věta 3.2 Nechť L je nekonečný regulární jazyk. Pak existuje celočíselná konstanta $p > 0$ taková, že platí:

$$w \in L \;\wedge\; |w| \geq p \quad\Rightarrow\quad \begin{array}{l} w = xyz \;\wedge \\ y \neq \varepsilon \;\wedge |xy| \leq p \;\wedge \\ xy^iz \in L \text{ pro } i \geq 0. \end{array}$$

❖ Poznámka: Neformálně řečeno Pumping lemma tvrdí, že v každé dostatečně dlouhé větě každého regulárního jazyka jsme schopni poblíž jejího začátku najít poměrně krátkou sekvenci, kterou je možné vypustit, resp. zopakovat libovolný počet krát, přičemž zůstáváme stále v rámci daného jazyka.

Věta 3.2 (z předchozí strany) Nechť L je nekonečný regulární jazyk. Pak existuje celočíselná konstanta $p > 0$ taková, že platí: $w \in L \;\wedge\; |w| \geq p \quad \Rightarrow \quad w = xyz \;\wedge$
$y \neq \varepsilon \;\wedge |xy| \leq p \;\wedge$
$xy^i z \in L$ pro $i \geq 0$.

Důkaz.
Nechť $L = L(M)$, $M = (Q, \Sigma, \delta, q_0, F)$ je konečný automat, kde $|Q| = n > 0$. Položme $p = n$. Je-li $w \in L$ a $|w| \geq n$, pak M přijme větu w „průchodem“ alespoň $n + 1$ konfiguracemi a tudíž alespoň dvě z nich obsahují stejný stav, tedy:

$$(q_0, w) = (q_0, xyz) \overset{*}{\vdash} (r, yz) \overset{k}{\vdash} (r, z) \overset{*}{\vdash} (q_F, \varepsilon),\; q_F \in F$$

pro nějaký stav $r \in Q$ a k takové, že $0 < k \leq n$. Dále je zřejmé, že k „zopakování“ stavu r dojde nejpozději po přečtení prvních n znaků vstupního řetězce a tudíž $|xy| \leq p$.

Pak ale existuje posloupnost konfigurací:

$$\begin{array}{ll} (q_0, xy^i z) & \overset{*}{\vdash} (r, y^i z) \\ & \overset{+}{\vdash} (r, y^{i-1} z) \\ & \vdots \\ & \overset{+}{\vdash} (r, z) \\ & \overset{*}{\vdash} (q_F, \varepsilon) \end{array}$$

Důkaz pokračuje dále.

Věta 3.2 (z předchozí strany) Nechť L je nekonečný regulární jazyk. Pak existuje celočíselná konstanta $p > 0$ taková, že platí: $w \in L \ \wedge \ |w| \geq p \quad \Rightarrow \quad w = xyz \ \wedge$
$y \neq \varepsilon \ \wedge |xy| \leq p \ \wedge$
$xy^i z \in L$ pro $i \geq 0$.

Pokračování důkazu. Z existence posloupnosti konfigurací:

$$\begin{array}{ll} (q_0, xy^i z) & \overset{*}{\vdash} (r, y^i z) \\ & \overset{+}{\vdash} (r, y^{i-1} z) \\ & \vdots \\ & \overset{+}{\vdash} (r, z) \\ & \overset{*}{\vdash} (q_F, \varepsilon) \end{array}$$

plyne $xy^i z \in L(M)$, a to nejen pro $i > 0$, ale i pro případ $i = 0$:

$$(q_0, xz) \overset{*}{\vdash} (r, z) \overset{*}{\vdash} (q_F, \varepsilon),\ q_F \in F$$

$\square$

❖ Věty 3.2 se často používá k důkazu, že daný jazyk není regulární.

Příklad 3.2 Dokažte, že jazyk $L = \{0^n 1^n \mid n \geq 1\}$ není regulární.

Důkaz:

❖ Předpokládejme, že L je regulární. Pak podle věty 3.2 existuje $p > 0$ takové, že každá věta $0^k 1^k$, kde $2k \geq p$, může být zapsána ve tvaru $xyz = 0^k 1^k$, $y \neq \varepsilon \ \wedge |xy| \leq p$, a platí $xy^i z \in L$ pro $i \geq 0$.

❖ Při hledání podřetězce y mohou nastat 3 možnosti:

$$0 \underbrace{0\,0 \ldots}_{y} \underbrace{.\,0\ \ 1\,1}_{y} \underbrace{1 \ldots}_{y} 1$$

$y \in \{0\}^+$	pak ale $xy^i z \notin L$ — nesouhlasí počet 0 a 1
$y \in \{1\}^+$	pak ale $xy^i z \notin L$ — nesouhlasí počet 0 a 1
$y \in \{0\}^+ \cdot \{1\}^+$	pak ale $xy^i z \notin L$ (všechny 0 nepředcházejí všechny 1)

❖ Řetězec y tudíž nelze vybrat, a proto $L \notin \mathcal{L}_3$.

❖ Nyní ukážeme, že vhodná volba slova může důkaz výrazně zjednodušit.

Alternativní důkaz příkladu 3.2:

❖ Předpokládejme, že L je regulární. Pak podle věty 3.2 existuje $p > 0$ takové, že věta $0^p 1^p$ může být zapsána ve tvaru $xyz = 0^p 1^p$, $y \neq \varepsilon \ \wedge |xy| \leq p$, a platí $xy^i z \in L$ pro $i \geq 0$.

❖ Při hledání podřetězce y může díky podmínce $|xy| \leq p$ nastat pouze jedna možnost:

$$0\,\underbrace{0\,0\,\ldots}_{y}\,.\,0\,1\,1\,1\ldots 1$$

$y \in \{0\}^+$ pak ale $xy^i z \notin L$ — nesouhlasí počet 0 a 1

❖ Řetězec y tudíž nelze vybrat, a proto $L \notin \mathcal{L}_3$.

$\square$

Příklad 3.3 Dokažte, že jazyk $L = \{a^q \mid q$ je prvočíslo $\}$ není regulární.

Důkaz:

- Předpokládejme, že L je regulární. Pak podle věty 3.2 existuje $p > 0$ takové, že každá věta $a^q \in L$, kde $q \geq p$, může být zapsána ve tvaru $a^q = xyz$, $y \neq \varepsilon$, $|xy| \leq p$, a platí $xy^iz \in L$ pro $i \geq 0$.

- Zvolme r prvočíslo větší než p.

- $a^r \in L$, $|a^r| = r$ a $r \geq p$, což implikuje $a^r = xyz$, $y \neq \varepsilon$, $|xy| \leq p$, a platí $xy^iz \in L$ pro $i \geq 0$.

- Nechť $|y| = k$ a zvolme $i = r + 1$.

- Pak ale $|xy^{r+1}z| = |xyz| + |y^r| = r + r.k = r.(k + 1)$, což však není prvočíslo, a tedy $xy^{r+1}z \notin L$. Spor.

$\square$

Myhill-Nerodova věta

Motivace

- Myhill-Nerodova věta
  - charakterizuje některé zásadní vztahy mezi konečnými automaty nad abecedou Σ a jistými ekvivalenčními relacemi nad řetězci ze Σ^*,
  - popisuje některé z nutných a postačujících podmínek pro to, aby daný jazyk byl jazykem regulárním (používá se často k důkazu neregularity jazyka),
  - poskytuje formální bázi pro elegantní důkaz existence unikátního (až na isomorfismus) minimálního DKA k danému regulárnímu jazyku.

Pravá kongruence a prefixová ekvivalence

❖ Zopakování: ekvivalence $\sim$ je binární relace, která je *reflexivní, symetrická a tranzitivní*. Index ekvivalence $\sim$ je počet tříd rozkladu $\Sigma^*/\sim$. Je-li těchto tříd nekonečně mnoho, definujeme index jako ∞.

Definice 3.1 Nechť Σ je abeceda a $\sim$ je ekvivalence na Σ^*. Ekvivalence $\sim$ je pravou kongruencí (je zprava invariantní), pokud pro každé $u, v, w \in \Sigma^*$ platí

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

Věta 3.3 Ekvivalence $\sim$ na Σ^* je pravá kongruence právě tehdy, když pro každé $u, v \in \Sigma^*$, $a \in \Sigma$ platí $u \sim v \Longrightarrow ua \sim va$.

Důkaz. „$\Rightarrow$“ je triviální, „$\Leftarrow$“ lze snadno ukázat indukcí nad délkou w. $\square$

Definice 3.2 Nechť L je libovolný (ne nutně regulární) jazyk nad abecedou Σ. Na množině Σ^* definujeme relaci $\sim_L$ zvanou prefixová ekvivalence pro L takto:

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

Myhill-Nerodova věta

Věta 3.4 Nechť L je jazyk nad Σ. Pak následující tvrzení jsou ekvivalentní:

1. L je jazyk přijímaný deterministickým konečným automatem.
2. L je sjednocením některých tříd rozkladu určeného pravou kongruencí na Σ^* s konečným indexem.
3. Relace $\sim_L$ má konečný index.

Důkaz. Dokážeme následující implikace:

- **1** $\Rightarrow$ **2**
- **2** $\Rightarrow$ **3**
- **3** $\Rightarrow$ **1**

Z definice ekvivalence $(a \Leftrightarrow b \overset{def}{\Longleftrightarrow} a \Rightarrow b \ \wedge \ b \Rightarrow a)$ a ze základní tautologie výrokové logiky $(a \Rightarrow b \ \wedge \ b \Rightarrow c) \Rightarrow (a \Rightarrow c)$ plyne tvrzení věty.

$\square$

Důkaz implikace 1 $\Rightarrow$ 2

❖ Je-li L přijímán DKA, pak L je sjednocením některých tříd rozkladu určeného pravou kongruencí na Σ^* s konečným indexem.

❖ Pro DKA $M = (Q, \Sigma, \delta, q_0, F)$ zaveďme zobecněnou přechodovou funkci $\hat{\delta} : Q \times \Sigma^* \rightarrow Q$ tak, že $\forall q_1, q_2 \in Q, w \in \Sigma^* : \hat{\delta}(q_1, w) = q_2 \Leftrightarrow (q_1, w) \overset{*}{\underset{M}{\vdash}} (q_2, \varepsilon)$.

Důkaz. Pro daný L přijímaný konečným automatem M zkonstruujeme $\sim$ s potřebnými vlastnostmi:

- Nechť $M = (Q, \Sigma, \delta, q_0, F)$ a δ je totální.
- Zvolíme $\sim$ jako binární relaci na Σ^* takovou, že $u \sim v \iff \hat{\delta}(q_0, u) = \hat{\delta}(q_0, v)$.
- Ukážeme, že $\sim$ má potřebné vlastnosti:
  - $\sim$ je *ekvivalence*: je reflexivní, symetrická a 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 některých tříd* $\Sigma^* \backslash \sim$: těch, které odpovídají F.

$\square$

Důkaz implikace 2 $\Rightarrow$ 3

❖ Existuje-li relace $\sim$ splňující podmínku 2, pak $\sim_L$ má konečný index.

Důkaz.

- Pro všechny $u, v \in \Sigma^*$ takové, že $u \sim v$, platí $u \sim_L v$:
  - Nechť $u \sim v$. Ukážeme, že také $u \sim_L v$, tj. $\forall w \in \Sigma^* : uw \in L \Leftrightarrow vw \in L$.
  - Víme, že $uw \sim vw$ a protože L je sjednocením některých tříd rozkladu $\Sigma^* \backslash \sim$, platí též $uw \in L \Leftrightarrow vw \in L$.
- Víme tedy, že $\sim \subseteq \sim_L$ (tj. $\sim_L$ je největší pravá kongruence s danými vlastnostmi).
- Každá třída $\sim$ je obsažena v nějaké třídě $\sim_L$.
- Index $\sim_L$ nemůže být větší než index $\sim$.
- $\sim$ má konečný index a tedy i $\sim_L$ má konečný index.

$\square$

Důkaz implikace 3 $\Rightarrow$ 1

❖ Má-li $\sim_L$ konečný index, pak L je přijímán nějakým konečným automatem.

Důkaz. Zkonstruujeme $M = (Q, \Sigma, \delta, q_0, F)$ přijímající L:

- $Q = \Sigma^* \backslash \sim_L$ (stavy jsou třídy rozkladu Σ^* relací $\sim_L$),
- $\forall u \in \Sigma^*, a \in \Sigma : \delta([u], a) = [ua]$,
- $q_0 = [\varepsilon]$,
- $F = \{[x] \mid x \in L\}$.

Uvedená konstrukce *je korektní*, tj. $L = L(M)$:

- Indukcí nad délkou slova v ukážeme, že $\forall v \in \Sigma^* : \hat{\delta}([\varepsilon], v) = [v]$.
- $v \in L \Longleftrightarrow [v] \in F \Longleftrightarrow \hat{\delta}([\varepsilon], v) \in F$.

$\square$

Důkaz neregularity pomocí M.-N. věty

Příklad 3.4 Dokažte, že jazyk $L = \{a^n b^n \mid n \geq 0\}$ není regulární.

Důkaz.

- Žádné řetězce $\varepsilon, a, a^2, a^3, ...$ nejsou $\sim_L$-ekvivalentní, protože $a^i b^i \in L$, ale $a^i b^j \notin L$ pro $i \neq j$.
- $\sim_L$ má tedy nekonečně mnoho tříd (neboli nekonečný index).
- Dle Myhill-Nerodovy věty tudíž nemůže být L přijímán žádným konečným automatem.

$\square$

M.-N. věta a minimalita DKA

Věta 3.5 (2. varianta Myhill-Nerodovy věty) Počet stavů libovolného minimálního DKA přijímajícího L je roven indexu $\sim_L$. (Takový DKA existuje právě tehdy, když je index $\sim_L$ konečný.)

Důkaz.

- Každý DKA (můžeme uvažovat DKA bez nedosažitelných stavů) určuje jistou pravou kongruenci s konečným indexem a naopak.
- Je-li L regulární, je $\sim_L$ největší pravou kongruencí s konečným indexem takovou, že L je sjednocením některých tříd příslušného rozkladu.
- Konečný automat, který odpovídá $\sim_L$ (viz důkaz 3 $\Rightarrow$ 1 Myhill-Nerodovy věty), je tedy minimální konečný automat přijímající L.

$\square$

Uzávěrové vlastnosti regulárních jazyků

Uzávěrové vlastnosti regulárních jazyků

Věta 3.6 Třída regulárních jazyků **je uzavřena** (mimo jiné) vzhledem k operacím:

$\cup$ (sjednocení),

$\cdot$ (konkatenace) a

* (iterace).

Důkaz. Plyne z definice regulárních množin a ekvivalence regulárních množin a regulárních jazyků. $\square$

Věta 3.7 Třída regulárních jazyků tvoří množinovou **Booleovu algebru**.

Důkaz.
Označme $\mathcal{L}_3$ třídu všech regulárních jazyků abecedou Σ. Ukážeme, že algebraická struktura $< \mathcal{L}_3, \cup, \cap, \ ', \Sigma^*, \emptyset >$ je Booleova algebra.

Důkaz pokračuje dále.

Pokračování důkazu.

1. Protože $\Sigma^* \in \mathcal{L}_3$ i $\emptyset \in \mathcal{L}_3$, jsou tyto jazyky zřejmě „jedničkou“, resp. „nulou“ uvažované algebry. Dále je zřejmá uzavřenost vůči $\cup$.
2. Dokážeme uzavřenost vzhledem ke komplementu nad abecedou Δ, $\Delta \subseteq \Sigma$. K jazyku L sestrojíme *úplně definovaný* KA M.

   $$M = (Q, \Delta, \delta, q_0, F)$$

   takový, že $L = L(M)$. Pak KA M'

   $$M' = (Q, \Delta, \delta, q_0, Q \setminus F)$$

   zřejmě přijímá jazyk $\Delta^* \setminus L$ (komplement L v Δ^*).
   Komplement vzhledem k Σ^*:

   $$\overline{L} = L(M') \cup \Sigma^*(\Sigma \setminus \Delta)\Sigma^*$$

   což je podle věty 3.3 regulární jazyk.

Důkaz pokračuje dále.

Pokračování důkazu.

3. Uzavřenost vzhledem k průniku plyne z de Morganových zákonů:

$$L_1 \cap L_2 = \overline{\overline{L_1 \cap L_2}} = \overline{\overline{L_1} \cup \overline{L_2}}$$

   a tedy $L_1, L_2 \in \mathcal{L}_3 \Rightarrow L_1 \cap L2 \in \mathcal{L}_3$.

$\square$

Věta 3.8 Nechť $L \in \mathcal{L}_3$ a nechť $L^R = \{w^R \mid w \in L\}$. Pak $L^R \in \mathcal{L}_3$.

Důkaz. Nechť M je KA takový, že $L = L(M)$. Sestrojíme M' tak, aby $L(M') = L^R$.
Popis konstrukce — na cvičení. $\square$

Rozhodnutelné problémy regulárních jazyků

Rozhodnutelné problémy v $\mathcal{L}_3$

Základní problémy:

- problém neprázdnosti: $L \neq \emptyset$?
- problém náležitosti: $w \in L$?
- problém ekvivalence: $L(G_1) = L(G_2)$?

Věta 3.9 Ve třídě $\mathcal{L}_3$ je rozhodnutelný problém **neprázdnosti** jazyka i problém **náležitosti** řetězce (do jazyka).

Důkaz.
K jazyku $L \in \mathcal{L}_3$ sestrojíme DKA M, $L = L(M)$:

$$M = (Q, \Sigma, \delta, q_0, F)$$

neprázdnost: $L(M) \neq \emptyset \iff \exists q \in Q : (q \in F \wedge q$ je dostupný z $q_0)$

náležitost: $w \in L \iff (q_0, w) \overset{*}{\vdash} (q, \varepsilon) \wedge q \in F$

□

Věta 3.10 Nechť $L_1 = L(G_1)$ a $L_2 = L(G_2)$ jsou dva jazyky generované regulárními gramatikami G_1 a G_2. Pak je rozhodnutelný problém **ekvivalence**, tj. $L(G_1) = L(G_2)$ nebo $L(G_1) \neq L(G_2)$.

Důkaz.
Nechť $M_1 = (Q_1, \Sigma_1, \delta_1, q_0^1, F_1)$, resp. $M_2 = (Q_2, \Sigma_2, \delta_2, q_0^2, F_2)$ jsou KA přijímající jazyky L_1, resp. L_2 takové, že $Q_1 \cap Q_2 = \emptyset$.

Vytvoříme konečný automat M takto:

$$M = (Q_1 \cup Q_2, \Sigma_1 \cup \Sigma_2, \delta_1 \cup \delta_2, q_0^1, F_1 \cup F_2)$$

a vypočítáme relaci $\equiv$ nerozlišitelnosti stavů z $Q_1 \cup Q_2$ pro automat M.
Pak

$$L(G_1) = L(G_2) \iff q_0^1 \equiv q_0^2$$

$\square$