Příklady pro cvičení 3. z IFJ:
Pumping lemma, uzávěrové vlastnosti

Slabé pumping lemma:
Nechť L je regulární jazyk. Potom existuje nějaká konstanta $k \geq 1$, že pro každý řetězce $z \in L$, kde $|z| \geq k$ platí, že jej můžeme rozdělit na podřetězce u, v, w, tedy $z = uvw$, které splňují následující podmínky: **1)** $v \neq \varepsilon$ **2)** $|v| \leq k$ **3)** $uv^mw \in L$ pro všechna $m \geq 0$.

Silné pumping lemma: *(učí se na přednáškách IFJ od roku 2005; silnější 2) podmínka)*
Nechť L je regulární jazyk. Potom existuje nějaká konstanta $k \geq 1$, že pro každý řetězce $z \in L$, kde $|z| \geq k$ platí, že jej můžeme rozdělit na podřetězce u, v, w, tedy $z = uvw$, které splňují následující podmínky: **1)** $v \neq \varepsilon$ **2)** $|uv| \leq k$ **3)** $uv^mw \in L$ pro všechna $m \geq 0$.

Příklad 1.
Pomocí pumping lemma dokažte, že jazyk $L_1 = \{a^nbba^n : n \geq 0\}$ není regulární.

Řešení (pomocí slabého PL):
1. Předpokládejme, že L_1 je regulární. Potom pro jazyk L_1 platí pumping lemma.
2. Nechť $\boldsymbol{k}$ je konstanta z pumping lemmy. Nyní vyberme „vhodný" řetězec z, který splňuje podmínky: **a)** $z \in L_1$ **b)** $|z| \geq \boldsymbol{k}$.
Položme tedy například $z = a^{\boldsymbol{k}}bba^{\boldsymbol{k}}$. Zřejmě $a^{\boldsymbol{k}}bba^{\boldsymbol{k}} \in L_1$ a $|a^{\boldsymbol{k}}bba^{\boldsymbol{k}}| = \boldsymbol{k} + 1 + 1 + \boldsymbol{k} = 2\boldsymbol{k} + 2 > \boldsymbol{k}$.
Obě podmínky **a)** i **b)** tedy jsou splněny.
3. Podle pumping lemma můžeme tedy řetězec $z = a^{\boldsymbol{k}}bba^{\boldsymbol{k}}$ „nějak" rozložit na tři podřetězce u, v, w, tak, aby platilo $z = uvw$. Nyní musíme projít **všechny možnosti** rozložení řetězce z a u každé z nich najít **aspoň jeden** případ, který vede ke sporu.

I. možnost:

$$\overbrace{aaa\ldots aa}^{k}bb\overbrace{aaa\ldots aaa}^{k}$$
$$\underbrace{aa}_{u}\boxed{\underbrace{a\ldots a}_{v}}\underbrace{aabbaaa\ldots aaa}_{w}$$

Formálně: $v = a^i$, $i \geq 1$; bb is substring of w

- Například pro řetězec $uv^0w = uw = \overbrace{aaaa}^{<k}bb\overbrace{aaa\ldots aaa}^{k} = \underbrace{aa}_{u}\underbrace{aabbaaa\ldots aaa}_{w}$
- Podle pumping lemma $uv^0w \in L_1$
- Podle definice jazyka L_1 ale $uv^0w \notin L_1$, protože nesouhlasí počet „a" (počet „a" před podřetězcem bb je menší než počet „a" za podřetězcem bb.
- **SPOR**

II. možnost:

$$\overbrace{aaa\ldots aa}^{k}bb\overbrace{aaa\ldots aaa}^{k}$$
$$\underbrace{aa}_{u}\boxed{\underbrace{a\ldots aaab}_{v}}\underbrace{baaa\ldots aaa}_{w}$$

Formálně: $v = a^ib$, $i \geq 0$;

- Například pro řetězec $uv^2w = uvvw =$ $\underbrace{aa}_{u}\underbrace{a\ldots aaab}_{v}\underbrace{a\ldots aaab}_{v}\underbrace{baaa\ldots aaa}_{w}$
- Podle pumping lemma $uv^2w \in L_1$
- Podle definice jazyka L_1 ale $uv^2w \notin L_1$, protože tento řetězec obsahuje tři symboly „b“.
- **SPOR**

III. možnost:

$$\overbrace{aaa\ldots aaa}^{k}bb\overbrace{aaa\ldots aaa}^{k}$$

$$\underbrace{aa}_{u}\underbrace{a\ldots aaabbaaa\ldots a}_{v}\underbrace{aa}_{w}$$

Formálně: $v = a^i bba^j$, $i, j \geq 0$;

- Například pro řetězec $uv^0w = uw =$ $\underbrace{aa}_{u}\underbrace{aa}_{w}$
- Podle pumping lemma $uv^0w \in L_1$
- Podle definice jazyka L_1 ale $uv^0w \notin L_1$, protože tento řetězec neobsahuje podřetězec bb.
- **SPOR**

IV. možnost:

$$\overbrace{aaa\ldots aaa}^{k}bb\overbrace{aaa\ldots aaa}^{k}$$

$$\underbrace{aaa\ldots aaab}_{u}\underbrace{baaa\ldots a}_{v}\underbrace{aa}_{w}$$

Formálně: $v = ba^i$, $i \geq 0$;

- Například pro řetězec $uv^2w = uvvw =$ $\underbrace{aaa\ldots aaab}_{u}\underbrace{baaa\ldots a}_{v}\underbrace{baaa\ldots a}_{v}\underbrace{aa}_{w}$
- Podle pumping lemma $uv^2w \in L_1$
- Podle definice jazyka L_1 ale $uv^2w \notin L_1$, protože tento řetězec obsahuje tři symboly „b“.
- **SPOR**

V. možnost:

$$\overbrace{aaa\ldots aaa}^{k}bb\overbrace{aaa\ldots aaa}^{k}$$

$$\underbrace{aaa\ldots aaabba}_{u}\underbrace{a\ldots a}_{v}\underbrace{aa}_{w}$$

Formálně: $v = a^i$, $i \geq 1$; bb is substring of u

- Například pro řetězec $uv^0w = uw =$ $\overbrace{aaa\ldots aaa}^{k}bb\overbrace{aaa}^{<k}$, $\underbrace{aaa\ldots aaabba}_{u}\underbrace{aa}_{w}$
- Podle pumping lemma $uv^0w \in L_1$
- Podle definice jazyka L_1 ale $uv^0w \notin L_1$, protože nesouhlasí počet „a“ (počet „a“ před podřetězcem bb je větší než počet „a“ za podřetězcem bb.
- **SPOR**

- Dokázali jsme sporem, že neexistuje žádné rozložení řetězce z na podřetětězce u, v, w, kde $z = uvw$ tak, aby pro všechna $m \geq 0$ platilo $uv^m w \in L_1$. Předpoklad, že L_1 je regulární, je nesprávný. (Jinak by toto rozložení podle pumping lemma muselo existovat)
  Proto tedy L_1 není regulární jazyk.

Poznámka: Použitím silného pumping lemma je možné provést pouze první možnost dekompozice, protože ostatní ihned porušují 2. podmínku: $|uv| \leq k$.

Příklad 2.
Pomocí výsledku z příkladu 1) a pomocí uzávěrových vlastností dokažte, že jazyk $L_2 = \{xy: x, y \in \{a, b\}^* \wedge y = \text{reversal}(x)\}$ není regulární. Neformálně, jazyk L_2 obsahuje všechny řetězce obsahující pouze symboly a, b, pro které platí, že přečtená druhá polovina řetězce od konce tvoří první polovinu řetězce.

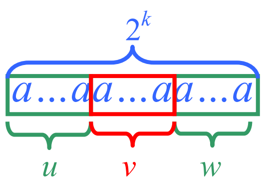
Řešení:
Důkaz sporem:
- Předpokládejme, že jazyk L_2 je regulární
- Regulární výraz $r = a^*bba^*$ zřejmě popisuje jistý regulární jazyk $L(r)$
- Regulární jazyky jsou uzavřeny vůči průniku, což znamená: Pro **libovolné** dva regulární jazyky L_a a L_b platí: $L_a \cap L_b$ je opět **jazyk regulární**.
- Konkrétně: L_2 je regulární (předpoklad), $L(r)$ je regulární (je popsán regulárním výrazem), jazyk $L_2 \cap L(r) = \{a^n bba^n: n \geq 0\}$ je tedy regulární.
  Ale v příkladu 1. jsme dokázali, že $\{a^n bba^n: n \geq 0\}$ není regulární $\Rightarrow$ **SPOR**.
- Předpoklad, že jazyk L_2 je regulární, je nesprávný.
- **Proto tedy L_2 není regulární jazyk.**

Poznámka: Nejsložitější částí příkladu je „vhodné“ zvolení regulárního výrazu r, tak aby po průniku s jazykem v zadání vznikl „jednodušší“ jazyk, u kterého lze efektivně pomocí pumping lemma dokázat, že není regulární.

Příklad 3.
Pomocí pumping lemma dokažte, že jazyk $L_3 = \{a^{2^n} : n \geq 0\}$ není regulární.

Řešení (Objevitelkou tohoto hezkého řešení je Jana Brychová, studentka 3BIT):
1. Předpokládejme, že L_3 je regulární. Potom pro jazyk L_3 platí pumping lemma.
2. Nechť $\boldsymbol{k}$ je konstanta z pumping lemmy. Nyní vyberme „vhodný“ řetězec z, který splňuje podmínky: **a)** $z \in L_3$ **b)** $|z| \geq \boldsymbol{k}$.
Položme tedy například $z = a^{2^k}$. Zřejmě $a^{2^k} \in L_3$ a $|a^{2^k}| = 2^k \geq k$. Obě podmínky **a)** i **b)** tedy jsou splněny.
3. Podle pumping lemma můžeme tedy řetězec $z = a^{2^k}$, „nějak“ rozložit na tři podřetězce u, v, w, tak, aby platilo $z = uvw$. Nyní musíme projít **všechny možnosti** rozložení řetězce z a u každé z nich najít **aspoň jeden** případ, který vede ke sporu.

I. možnost (jediná):



Formálně: $v = a^i$, $i \geq 1$;

Zřejmě $|uw| \geq 1$, neboť $|uw| = 0$ by implikovalo, že $|v| = 2^k$, což ihned dává spor s podmínkou $|v| \leq k$ u slabého pumping lemma nebo s podmínkou $|uv| \leq k$ u silného pumping lemma.

- $|z| = |uvw| = |uw| + |v| = 2^k$. Vynásobením této rovnice číslem 2 obdržíme: $2|uw| + 2|v| = 2.2^k$, tedy: $2|uw| + 2|v| = 2^{k+1} \ldots [1]$
- Podle pumping lemma $uv^2w \in L_3$, tedy $|uv^2w| = |uw| + 2|v| = 2^n \ldots [2]$ kde $\underline{n > k}$, neboť $|uv^2w| > |uvw|$.
- Odečtením [1] – [2] dostáváme:
  $2|uw| + 2|v| - (|uw| + 2|v|) = 2^{k+1} - 2^n$, tedy: $|uw| = 2^{k+1} - 2^n$
  Protože $|uw| \geq 1$, musí platit $2^{k+1} > 2^n$, čili $\underline{k+1 > n}$

Spor! (neexistuje přirozené číslo *n* mezi čísly *k* a *k*+1)

- Dokázali jsme sporem, že neexistuje žádné rozložení řetězce z na podřetětězce u, v, w, kde $z = uvw$ tak, aby pro všechna $m \geq 0$ platilo $uv^mw \in L_3$. Předpoklad, že L_3 je regulární, je nesprávný (Jinak by toto rozložení podle pumping lemma muselo existovat).
  Proto tedy L_3 není regulární jazyk.

- Dále jsme dokázali, že máme na fakultě chytré holky ;-)

Příklad 4.
Pomocí výsledku z příkladu 3) a pomocí uzávěrových vlastností dokažte, že jazyk $L_4 = \{x: x \in \{a, b, c\}^* \wedge |x| = 2^n, n \geq 0\}$ není regulární. Neformálně, jazyk L_4 obsahuje všechny řetězce obsahující pouze symboly a, b, c, pro které platí, že jejich délka je ve tvaru 2^n, $n \geq 0$.

Řešení:
Důkaz sporem:
- Předpokládejme, že jazyk L_4 je regulární
- Regulární výraz $r = a^*$ zřejmě popisuje jistý regulární jazyk $L(r)$
- Regulární jazyky jsou uzavřeny vůči průniku, což znamená: Pro **libovolné** dva regulární jazyky L_a a L_b platí: $L_a \cap L_b$ je opět **jazyk regulární**.
- Konkrétně: L_4 je regulární (předpoklad), $L(r)$ je regulární (je popsán regulárním výrazem), jazyk $L_4 \cap L(r) = \{ a^{2^n} : n \geq 0\}$ je tedy regulární.
  Ale v příkladu 3. jsme dokázali, že $\{ a^{2^n} : n \geq 0\}$ není regulární $\Rightarrow$ **SPOR**.
- Předpoklad, že jazyk L_4 je regulární, je nesprávný.
- **Proto tedy L_4 není regulární jazyk.**

Příklad 5.
Pomocí pumping lemma dokažte, že jazyk $L_5 = \{a^n: n \text{ je prvočíslo}\}$ není regulární.

Řešení:
1. Předpokládejme, že L_5 je regulární. Potom pro jazyk L_5 platí pumping lemma.
2. Nechť $\boldsymbol{k}$ je konstanta z pumping lemmy. Nyní vyberme „vhodný“ řetězec z, který splňuje podmínky: **a)** $z \in L_5$ **b)** $|z| \geq \boldsymbol{k}$.
Položme tedy například $z = a^p$, kde p je první prvočíslo splňující podmínku $p > \boldsymbol{k}$.
Zřejmě $a^p \in L_5$ a $|a^p| = \boldsymbol{p} \geq \boldsymbol{k}$. Obě podmínky **a)** i **b)** tedy jsou splněny.
3. Podle pumping lemma můžeme tedy řetězec $z = a^p$, kde p je prvočíslo, „nějak“ rozložit na tři podřetězce u, v, w, tak, aby platilo $z = uvw$. Nyní musíme projít **všechny možnosti** rozložení řetězce z a u každé z nich najít **aspoň jeden** případ, který vede ke sporu.

Mohli bychom postupovat zcela obecně a vzít pouze jedinou možnost (jako v příkladu 3.):

$$\overbrace{\underbrace{a\ldots a}_{u}\underbrace{a\ldots a}_{v}\underbrace{a\ldots a}_{w}}^{p}$$

Formálně: $v = a^i$, $i \geq 1$;

Takto důkaz sice je možné provést, ale je příliš komplikovaný, proto tento případ rozdělme na následující tři možnosti:

I. možnost:

$$\overbrace{\underbrace{aaaa\ \ldots\ aaaa}_{v}}^{p}$$

Formálně: $v = a^i$, $i \geq 1$;
$u, w = \varepsilon$, tedy: $|uw| = 0$

- Například pro řetězec $uv^0w = uw = \underset{uw}{\boxed{}\boxed{}} = \varepsilon$:
- Podle pumping lemma $uv^0w \in L_5$
- $uv^0w = \varepsilon$, $|uv^0w| = 0$. Podle definice jazyk L_5 obsahuje pouze řetězce, které mají délku prvočísla. 0 ale není prvočíslo, tedy $uv^0w \notin L_5$.
- **SPOR**

II. možnost:

$$\overbrace{a\,\underbrace{aaa\ \ldots\ aaaa}_{v}}^{p}$$

nebo

$$\overbrace{\underbrace{aaaa\ \ldots\ aaa}_{v}\,a}^{p}$$

Formálně: $v = a^i$, $i \geq 1$;
$u = a$, $w = \varepsilon$ nebo $u = \varepsilon$, $w = a$,
tedy: $|uw| = 1$

- Například pro řetězec $uv^0w = uw = \underset{u\ \ w}{\boxed{a}\boxed{}}$ nebo $\underset{u\ \ w}{\boxed{}\boxed{a}} = a$:
- Podle pumping lemma $uv^0w \in L_5$
- $uv^0w = a$, $|uv^0w| = 1$. Podle definice jazyk L_5 obsahuje pouze řetězce, které mají délku prvočísla. 1 ale není prvočíslo, tedy $uv^0w \notin L_5$.
- **SPOR**

III. možnost (všechny ostatní případy):

$$\overbrace{\underbrace{a\ldots a}_{u}\underbrace{a\ldots a}_{v}\underbrace{a\ldots a}_{w}}^{p}$$

, kde $|uw| \geq 2$,

Formálně: $v = a^i$, $i \geq 1$;
$|uw| \geq 2$

Položme $m = |uw|$.

- Podle pumping lemma $uv^mw \in L_5$. (Platí pro libovolné $m \geq 0$, tedy i pro $m = |uw|$).
- $|uv^mw| = \underbrace{|uw|}_{m} + |v^m| = m + m|v| = \underbrace{m}_{= |uw| \geq 2} \,.\, \underbrace{(1 + |v|)}_{\geq 2,\ \text{neboť } |v| \geq 1}$

Je tedy vidět, že $|uv^m w|$ můžeme rozložit na součin dvou přirozených čísel, z nichž obě jsou větší nebo rovny číslu 2. Zřejmě tedy $|uv^m w|$ není prvočíslo, neboť prvočísla mají dělitele pouze 1 a sama sebe. Podle definice jazyk L_5 obsahuje pouze řetězce, které mají délku prvočísla. $|uv^m w|$ ale není prvočíslo, tedy $uv^m w \notin L_5$.

- **SPOR**
- Dokázali jsme sporem, že neexistuje žádné rozložení řetězce z na podřetětězce u, v, w, kde $z = uvw$ tak, aby pro všechna $m \geq 0$ platilo $uv^m w \in L_5$. Předpoklad, že L_5 je regulární, je nesprávný (Jinak by toto rozložení podle pumping lemma muselo existovat).
  Proto tedy L_5 není regulární jazyk.

Příklad 6.
Přednášky ukázaly pomocí pumping lemma, že jazyk $\{a^n b^n : n \geq 0\}$ není regulární. S využitím tohoto poznatku a uzávěrových vlastností dokažte, že ani jazyk $L_6 = \{x : x \in \{a, b\}^*, \#_a x = \#_b x\}$ není regulární.

Poznámka: $\#_a x$ obecně znamená počet výskytů symbolů a v řetězci x, tedy podmínka $\#_a x = \#_b x$ popisuje skutečnost, že počty výskytů symbolů a i b jsou v řetězci x stejné.

Řešení:
Důkaz sporem:

- Předpokládejme, že jazyk L_6 je regulární
- Regulární výraz $r = a^* b^*$ zřejmě popisuje jistý regulární jazyk $L(r)$
- Regulární jazyky jsou uzavřeny vůči průniku, což znamená: Pro **libovolné** dva regulární jazyky L_a a L_b platí: $L_a \cap L_b$ je opět **jazyk regulární**.
- Konkrétně: L_6 je regulární (předpoklad), $L(r)$ je regulární (je popsán regulárním výrazem).
  Jazyk $L_6 \cap L(r) = \{a^n b^n : n \geq 0\}$ je tedy regulární.
  Ale v přednášky ukázali, že $\{a^n b^n : n \geq 0\}$ není regulární $\Rightarrow$ **SPOR**.
- Předpoklad, že jazyk L_6 je regulární, je nesprávný.
- **Proto tedy L_6 není regulární jazyk.**

Příklad 7.
Nechť L_7 je jazyk nad abecedou Σ = { „(“ , „)“ } obsahující všechny řetězce, které vzniknou z aritmetických výrazů vypuštěním všech symbolů kromě závorek. Například ( )(( )( )) $\in L_7$, neboť tento řetězec vznikl například z aritmetického výrazu $(a + b).((a + b) - (c + d))$.

a) Dokažte, že tento jazyk není regulární
b) Ukažte, že tento jazyk splňuje podmínky slabého pumping lemma pro regulární jazyky
c) Jaká značná nepříjemnost vyplývá z podmínek a) a b)?

Řešení:
a) Důkaz sporem:

- Předpokládejme, že jazyk L_7 je regulární
- Regulární výraz r = „(“* „)“* zřejmě popisuje jistý regulární jazyk $L(r)$
- Regulární jazyky jsou uzavřeny vůči průniku, což znamená: Pro **libovolné** dva regulární jazyky L_a a L_b platí: $L_a \cap L_b$ je opět **jazyk regulární**.
- Konkrétně: L_7 je regulární (předpoklad), $L(r)$ je regulární (je popsán regulárním výrazem).
  Jazyk $L_7 \cap L(r)$ = {„(“^{n} „)“^{n}: $n \geq 0$} je tedy regulární.

- Přednášky ale ukázaly, že $\{a^n b^n: n \geq 0\}$ není regulární, tedy ani {„(“^{n}„)“^{n}: $n \geq 0$} není regulární $\Rightarrow$ **SPOR**.
- Předpoklad, že jazyk L_7 je regulární, je nesprávný.
- **Proto tedy L_7 není regulární jazyk.**

b) Dokažme, že pro jazyk L_7 platí slabší pumping lemma: Zvolme konstantu $k = 2$. Zřejmě musí každý „závorkový“ řetězec patřící do jazyka L_7 délky větší nebo rovny číslu 2 obsahovat podřetězec „( )“ jakožto nějaký pár závorek, které již v sobě jiné závorky neobsahují. Proveďme následující dekompozici daného řetězce na podřetězce ***u***, ***v***, ***w*** tak, aby řetězec v = „( )“:

$$\underbrace{(\sim ? \sim}_{u}\underbrace{(\,)}_{v}\underbrace{\sim ? \sim)}_{w}$$

Zřejmě i řetězce:

$$\underbrace{(\sim ? \sim}_{u}\underbrace{\sim ? \sim)}_{w},\ \underbrace{(\sim ? \sim}_{u}\underbrace{(\,)}_{v}\underbrace{(\,)}_{v}\underbrace{\sim ? \sim)}_{w},\ \underbrace{(\sim ? \sim}_{u}\underbrace{(\,)}_{v}\underbrace{(\,)}_{v}\underbrace{(\,)}_{v}\underbrace{\sim ? \sim)}_{w},\ \ldots,\ \underbrace{(\sim ? \sim}_{u}\underbrace{(\,)}_{v}\underbrace{(\,)}_{v}\ldots\underbrace{(\,)}_{v}\underbrace{\sim ? \sim)}_{w},\ \ldots$$

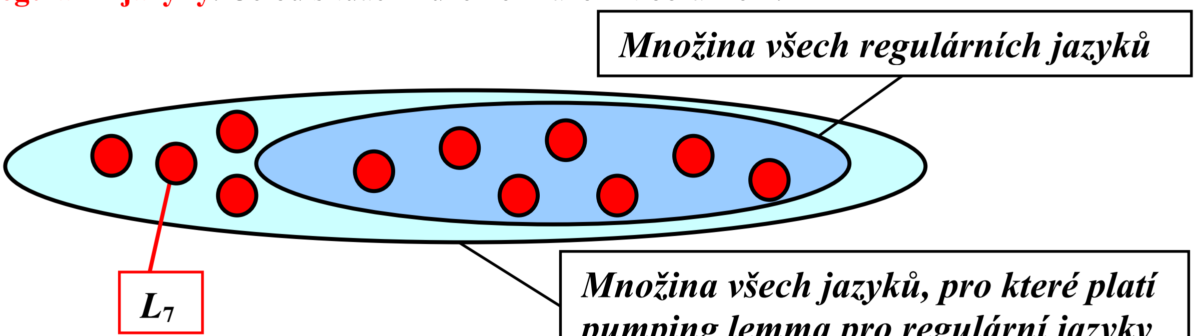
jsou obsaženy v jazyce L_7, tedy platí:

1) $v \neq \varepsilon$
2) $|v| \leq k$ } neboť $|v| = 2$, $k = 2$
3) $uv^m w \in L_7$ pro všechna $m \geq 0$.

- Pro daný jazyk L_7 tedy platí pumping lemma pro regulární jazyky.

c) Zřejmě pro každý regulární jazyk platí pumping lemma pro regulární jazyky (viz. přednášky). Jazyk L_7 není regulární (viz. část a) a přesto pro něj pumping lemma pro regulární jazyky platí (viz. část b).

- **Existují tedy jazyky, které nejsou regulární, a přesto pro ně platí pumping lemma pro regulární jazyky**. Celou situaci můžeme znázornit obrázkem:



Nepříjemný důsledek: Pumping lemma pro regulární jazyky tedy pouze říká:

- **POKUD** L je regulární jazyk, **POTOM** pro něj platí podmínky PL

A NE:

- ~~**POKUD** pro daný jazyk L platí podmínky PL **POTOM** L je regulární jazyk~~

Například jazyk L_7 splňuje podmínky pumping lemma pro regulární jazyky a přesto tento jazyk není regulární !!!

Závěr:
Pomocí pumping lemma nemůžeme nikdy dokázat, že daný jazyk je regulární!!!