# Poznámky k cvičeniu č. 8

Peter Kostolányi

15. novembra 2017

## Regulárne jazyky

Ako bolo dokázané na prednáške, deterministické konečné automaty, nedeterministické konečné automaty a regulárne gramatiky sú rovnako silné modely opisujúce rovnakú triedu jazykov. Jazyky z tejto triedy nazývame *regulárne.*

**Definícia 1.** *Regulárny jazyk* je jazyk $L$, pre ktorý existuje deterministický konečný automat $A$ taký, že $L(A) = L$. Triedu všetkých regulárnych jazykov označujeme symbolom $\mathscr{R}$.

**Veta 1.** *Nech $L$ je jazyk. Nasledujúce tvrdenia sú ekvivalentné:*

(*i*) *Jazyk $L$ je regulárny.*

(*ii*) *Existuje nedeterministický konečný automat $A$ taký, že $L(A) = L$.*

(*iii*) *Existuje regulárna gramatika $G$ taká, že $L(G) = L$.*

## Pumpovacia lema pre regulárne jazyky

Nech $L \in \mathscr{R}$ je *ľubovoľný* regulárny jazyk. V takom prípade *existuje* deterministický konečný automat $A = (K, \Sigma, \delta, q_0, F)$ taký, že $L(A) = L$. Ten má nejaký počet stavov $|K| =: p = q$.[1] Nech $w \in L$ je *ľubovoľné* slovo z jazyka $L$ také, že $|w| \geq p$. Z Dirichletovho princípu vyplýva, že počas výpočtu automatu $A$ na tomto slove sa niektorý jeho stav $q_{op}$ musí „vyskytnúť“ aspoň dvakrát. *Existujú* teda slová $u, v \in \Sigma^*$ a $x \in \Sigma^+$ také, že $w = uxv$ a

$$(q_0, uxv) \vdash^* (q_{op}, xv) \vdash^+ (q_{op}, v) \vdash^* (q_{fin}, \varepsilon),$$

kde $q_{fin} \in F$ je akceptačný stav. Nech teraz $i \in \mathbb{N}$ je ľubovoľné prirodzené číslo. Ľahko možno vidieť, že existuje akceptačný výpočet automatu $A$ na slove $ux^iv$: stačí namiesto jedného „cyklu“ $(q_{op}, x) \vdash^+ (q_{op}, \varepsilon)$ vykonať takýchto „cyklov“ niekoľko. Inými slovami: platí

$$(q_0, ux^iv) \vdash^* (q_{op}, x^iv) \vdash^+ (q_{op}, x^{i-1}v) \vdash^+ \ldots \vdash^+ (q_{op}, v) \vdash^* (q_{fin}, \varepsilon).$$

V prípade, že $i = 0$, zodpovedá uvedený zápis „vynechaniu cyklu“. Je navyše zrejmé, že prvý opakovaný výskyt nejakého stavu musí prísť po najviac $q$ krokoch výpočtu, takže možno pridať obmedzenie $|ux| \leq q$.

Pomocou takýchto jednoduchých úvah možno odvodiť pumpovaciu lemu pre regulárne jazyky, pričom správne poradie kvantifikátorov je dané poradím slov, ktoré sme zvýrazňovali italikou.

**Veta 2.** *Nech $L \in \mathscr{R}$ je regulárny jazyk. Potom existujú $p, q \in \mathbb{N}$ také, že pre všetky $w \in L$, $|w| \geq p$, existujú slová $u, x, v \in \Sigma_L^*$ také, že platí:*

(*i*) $w = uxv$,

(*ii*) $|ux| \leq q$,

(*iii*) $|x| \geq 1$,

(*iv*) $\forall i \in \mathbb{N} : ux^iv \in L$.

Je odporúčané pumpovacej leme skutočne porozumieť a namiesto učenia sa jej naspamäť ju radšej vždy „na počkanie“ odvodiť. Priestor na chyby (predovšetkým v poradí kvantifikátorov) je v takom prípade omnoho menší.

[1] V pumpovacej leme sa táto hodnota zvykne označovať $p$ resp. $q$. Ide tu o tradičné označenie, ktoré občas môže narobiť jemné problémy so zaužívanými konvenciami označovania stavov.

**Dokazovanie negatívnych výsledkov o regularite jazykov**

Keďže je pumpovacia lema sformulovaná ako tvrdenie platné pre každý regulárny jazyk, platnosť podmienok z nej vyplývajúcich je pre jazyk $L$ *nutnou podmienkou* jeho regularity. Ak teda nejaký jazyk nespĺňa podmienky vyplývajúce z pumpovacej lemy, nemôže byť regulárny. Táto skutočnosť je základom jednej z najefektívnejších metód dokazovania negatívnych výsledkov o regularite jazykov, ktorú v nasledujúcom demonštrujeme na príklade jazyka všetkých palindrómov nad abecedou $\{a,b\}$, o ktorom je intuitívne zrejmé, že nie je regulárny.

**Úloha 1.** Dokážte, že jazyk $L = \{w \in \{a,b\}^* \mid w = w^R\}$ nie je regulárny.

*Riešenie.* Sporom, nech $L \in \mathscr{R}$. Jazyku $L$ potom podľa pumpovacej lemy prislúchajú čísla $p, q \in \mathbb{N}$. Vezmime slovo $w = a^{p+q}ba^{p+q}$ – zjavne platí $w \in L$ a $|w| \geq p$. Preto existujú slová $u, x, v \in \{a,b\}^*$, pre ktoré sú splnené podmienky $(i)$ až $(iv)$ pumpovacej lemy.

Z podmienky $(i)$ máme $w = uxv$. Z podmienok $(ii)$ a $(iii)$ vyplýva, že existujú $r, s \in \mathbb{N}$ také, že $s \geq 1$, $u = a^r$, $x = a^s$, a $v = a^{p+q-s-r}ba^{p+q}$.

Nakoniec, z podmienky $(iv)$ pumpovacej lemy pre $i = 2$ vyplýva, že slovo

$$ux^2v = a^r a^{2s} a^{p+q-s-r} ba^{p+q} = a^{p+q+s} ba^{p+q}$$

patrí do jazyka $L$. To je ale spor, pretože $s \geq 1$. □

Pri nasledujúcej „písomkovej“ úlohe uvádzame okrem jej správneho riešenia aj nesprávne riešenie, ktorého variácie sa vyskytovali pomerne často.

**Úloha 2.** Zistite, či je jazyk $L = \{a^{2^n} \mid n \in \mathbb{N}\}$ regulárny. Svoje tvrdenie dokážte.

***Nesprávne*** *riešenie.* Dokážeme, že jazyk $L$ nie je regulárny. Sporom, nech $L \in \mathscr{R}$. Nech $p, q \in \mathbb{N}$ sú konštanty zodpovedajúce jazyku $L$ podľa pumpovacej lemy. Bez ujmy na všeobecnosti môžeme predpokladať, že $p + q \geq 1$. Vezmime slovo $w = a^{2^{p+q}}$. Potom $w = uxv$, kde $u = a$, $x = a$ a $v = a^{2^{p+q}-2}$. Nech $i = 2$. Potom z podmienky $(iv)$ pumpovacej lemy vyplýva $ux^2v = a^{2^{p+q}+1} \in L$, čo je spor, pretože $2^{p+q} + 1$ nemôže byť mocninou čísla 2.

Toto riešenie **je chybné**, pretože pumpovacia lema nám nedovoľuje zvoliť si slová $u, x, v$ úplne ľubovoľne. Jediné, čo nám dovoľuje o týchto slovách predpokladať, sú podmienky $(i)$ až $(iii)$. □

***Správne*** *riešenie.* Dokážeme, že jazyk $L$ nie je regulárny. Sporom, nech $L \in \mathscr{R}$. Nech $p, q \in \mathbb{N}$ sú konštanty zodpovedajúce jazyku $L$ podľa pumpovacej lemy. Vezmime teraz ľubovoľné slovo $w = a^{2^m}$ také, že $2^m > p + q$ – očividne $w \in L$ a $|w| \geq p$. Potom existujú slová $u, x, v$, pre ktoré sú splnené podmienky $(i)$ až $(iv)$ pumpovacej lemy.

Z podmienky $(i)$ máme $w = uxv$. Z podmienok $(ii)$ a $(iii)$ vyplýva, že existujú čísla $r, s \in \mathbb{N}$ také, že $s \geq 1$, $r + s \leq q$, $u = a^r$, $x = a^s$ a $v = a^{2^m - r - s}$. Z podmienky $(iv)$ pumpovacej lemy potom vyplýva $ux^2v = a^{2^m + s} \in L$. Keďže ale $s \geq 1$ a $s \leq r + s \leq q < 2^m$, platí $2^m < 2^m + s < 2^{m+1}$, a teda $2^m + s$ nie je mocnina dvoch, čo je spor. □

*Poznámka* 1. Vyvarovať sa chýb podobných tej vyššie je možné len pri správnom pochopení toho, čo sa pri dôkazoch s použitím pumpovacej lemy deje.

Typický takýto dôkaz je dôkazom sporom – za účelom sporu predpokladáme, že jazyk $L$, o ktorom chceme ukázať, že nie je regulárny, regulárny je. Pumpovacia lema je zárukou, že za tohto predpokladu platí pre $L$ určité tvrdenie. K sporu chceme prísť tak, že dokážeme negáciu tohto tvrdenia – teda, že *pre všetky* $p, q \in \mathbb{N}$ *existuje* $w \in L$ s $|w| \geq p$ také, že *pre žiadne* $u, x, v \in \Sigma_L^*$ nemôže súčasne platiť $(i)$ až $(iv)$.

To znamená, že kým s konštantami $p, q$ musíme pracovať *vo všeobecnosti*, slovo $w$ si môžeme zvoliť ako *ľubovoľné* slovo z jazyka $L$ dĺžky aspoň $p$. Slová $u, x, v$ ale opäť musíme uvažovať vo všeobecnosti – potrebujeme totiž ukázať, že pre žiadne prípustné $u, x, v$ nemôžu byť súčasne splnené všetky podmienky $(i)$ až $(iv)$. To ale väčšinou dokazujeme tak, že predpokladáme platnosť podmienok $(i)$ až $(iii)$ a dokazujeme, že neplatí $(iv)$. Slová $u, x, v$ teda predsa len nemusíme uvažovať úplne ľubovoľné – môžeme predpokladať, že sú pre ne splnené podmienky $(i)$ až $(iii)$. Nič viac ale predpokladať nemôžeme.

**Uzáverové vlastnosti triedy regulárnych jazykov**

Trieda jazykov $\mathscr{L}$ je uzavretá na nejakú $k$-árnu operáciu $Op$, ak $Op(L_1, \dots, L_k) \in \mathscr{L}$ kedykoľvek $L_1, \dots, L_k \in \mathscr{L}$. Teda napríklad trieda regulárnych jazykov $\mathscr{R}$ je uzavretá na zjednotenie, pretože pre všetky $L_1, L_2 \in \mathscr{R}$ platí $L_1 \cup L_2 \in \mathscr{R}$.

Dokázať uzavretosť triedy regulárnych jazykov na $k$-árnu operáciu operáciu $Op$ teda znamená pre všetky regulárne jazyky $L_1, \dots, L_k$ dokázať, že aj jazyk $Op(L_1, \dots, L_k)$ je regulárny. Regulárnosť jazyka $Op(L_1, \dots, L_k)$ sa dá dokázať napríklad všeobecnou konštrukciou DKA, NKA alebo regulárnej gramatiky na základe DKA, NKA alebo regulárnych gramatík, ktorých existenciu predpokladáme pre jazyky $L_1, \dots, L_k$. Vyvrátiť uzavretosť triedy $\mathscr{R}$ na operáciu $Op$ naopak znamená nájsť konkrétny príklad jazykov $L_1, \dots, L_k \in \mathscr{R}$ takých, že $Op(L_1, \dots, L_k) \notin \mathscr{R}$.

**Úloha 3.** Zistite, či je trieda $\mathscr{R}$ uzavretá na unárnu operáciu *rotácie*, definovanú pre všetky jazyky $L$ nasledovne:

$$\mathrm{rot}(L) = \{vu \mid u, v \in \Sigma_L^*;\ uv \in L\}.$$

Svoje tvrdenie dokážte.

*Riešenie.* Dokážeme, že trieda $\mathscr{R}$ *je uzavretá* na rotáciu. Nech $L \in \mathscr{R}$ je regulárny jazyk. Potom existuje deterministický konečný automat $A$ taký, že $L(A) = L$. Zostrojíme nedeterministický konečný automat $A'$ taký, že $L(A') = \mathrm{rot}(L)$.

Automat $A'$ dostane vstup $w$, ktorý má akceptovať v prípade, že existujú slová $u, v$, pre ktoré platí $w = vu$ a zároveň $uv \in L$. Ak $uv \in L$, musí existovať akceptačný výpočet automatu $A$ na slove $uv$ – tento výpočet bude „základom“ pre výpočet automatu $A'$. Automat $A'$ teda najprv „nedeterministicky uhádne“ stav $q$, v ktorom automat $A$ dočíta slovo $u$ – správnosť tohto „tipu“ overí na konci výpočtu. Následne spustí simuláciu automatu $A$ na vstupe $w$ – ktoré má v prípade akceptácie zodpovedať slovu $vu$ – so začiatkom simulácie v stave $q$. Automat $A'$ bude simulovať automat $A$, až kým sa „nedeterministicky rozhodne“, že prečítal slovo $v$.

Ak táto simulácia skončila v neakceptačnom stave automatu $A$, buď na začiatku nešlo o správny „nedeterministický tip“ stavu $q$, alebo automat $A$ slovo $uv$ neakceptuje. V oboch prípadoch automat $A'$ svoj vstup zamietne. Ak simulácia naopak skončila v akceptačnom stave, automat $A'$ overí svoj „nedeterministický tip“ stavu $q$. To znamená, že na zvyšku svojho vstupu – ten má zodpovedať slovu $u$ – spustí simuláciu automatu $A$ so začiatkom simulácie v stave $q_0$ a overí, či po jeho dočítaní bude automat $A$ naozaj v stave $q$. Ak áno, automat $A'$ svoj vstup akceptuje.

Nech teda $A = (K, \Sigma, \delta, q_0, F)$. Potom $A' = (K', \Sigma', \delta', q_0', F')$, kde $K' = \{q_0'\} \cup K \times \{1, 2\} \times K$, $\Sigma' = \Sigma$, pre prechodovú funkciu $\delta'$ platí

$$\begin{aligned}
\delta'(q_0, \varepsilon) &= \{(q, 1, q) \mid q \in K\},\\
\forall p, q \in K\ \forall c \in \Sigma : \delta'((p, 1, q), c) &= \{(\delta(p, c), 1, q)\},\\
\forall p \in F\ \forall q \in K : \delta'((p, 1, q), \varepsilon) &= \{(q_0, 2, q)\},\\
\forall p, q \in K\ \forall c \in \Sigma : \delta'((p, 2, q), c) &= \{(\delta(p, c), 2, q)\}
\end{aligned}$$

(ostatné výstupy prechodovej funkcie sú prázdne množiny) a množina akceptačných stavov $F'$ je

$$F' = \{(q, 2, q) \mid q \in K\}.$$

Dôkaz správnosti tejto konštrukcie robiť nebudeme. □

**Úloha 4.** Zistite, či je trieda $\mathscr{R}$ uzavretá na (vymyslenú) binárnu operáciu „zdvojenej“ iterácie, definovanú pre všetky jazyky $L_1, L_2$ nasledovne:

$$2\text{-it}(L_1, L_2) = \bigcup_{i=0}^{\infty} \left(L_1^i \cdot L_2^i\right).$$

Svoje tvrdenie dokážte.

*Riešenie.* Dokážeme, že trieda $\mathscr{R}$ *nie je uzavretá* na túto operáciu. Uvažujme napríklad regulárne jazyky $L_1 = \{a\}$ a $L_2 = \{b\}$. Potom zrejme $2\text{-it}(L_1, L_2) = \{a^n b^n \mid n \in \mathbb{N}\}$, čo je jazyk, o ktorom vieme, že nie je regulárny. □