Léma 3.12

Majme konečný automat $A = (Q, \Sigma, \delta, q_0, F)$ Nech pre $x, y \in \Sigma^*, x \neq y$:

$$\hat{\delta}(q_0, x) = \hat{\delta}(q_0, y) = p \tag{1}$$

pre nejaké $p \in Q$.

Potom pre ľubovoľné $z \in \Sigma^*$ existuje $r \in Q$ také, že $xz \in KL[r]$ aj $yz \in KL[r]$ a platí, že:

$$xz \in L(A) \Leftrightarrow yz \in L(A) \tag{2}$$

Dôkaz neregulárnosti pomocou Lémy 3.12

- Dokazujeme sporom, takže predpokladáme, že jazyk L je regulárny a teda existuje DKA A, ktorý ho rozoznáva $L = L(A)$.
- Ak máme konečný automat s n stavmi, tak ak budeme mať čo i len o jedno slovo viac, ako je stavov $(n + 1)$, musia výpočty na aspoň dvoch slovách skončiť v rovnakom stave (Pigeonhole principle).
- Zvolíme si nejakú množinu slov nad abecedou (nemusia byť z jazyka, dokonca často je jednoduchšie, aby z jazyka neboli), ktorých je dostatočne veľa (aspoň $n + 1$ pre ľubovoľné n), takže napr. 0^n, a^{2k}... Treba vysvetliť/ukázať, prečo je ich určite viac ako stavov automatu. Z nich vyberiete dve slová. Pozor, nie ľubovoľné! Treba si uvedomiť, že vy neviete, ktoré konkrétne dve slová skončia v rovnakom stave, ale viete, že ak ste vytvorili aspoň $n + 1$ slov, tak to mohlo byť ľubovolné i-te a j-te. Preto *nevhodné slová* sú také, kde používate konštanty a taktiež také, kde jedno závisí od druhého napr. $1^i, 1^{2i}$ alebo a^5, a^{5+k}.
- Z Lémy 3.12 vieme, že ak výpočet na slovách u a v skončí v rovnakom stave q, tak aj výpočty na ux a vx skončia v jednom stave p, čiže slovo ux patrí do jazyka L práve vtedy keď aj slovo vx.
- Slovo x je preto vhodné vybrať tak, aby jedno slovo z dvojice ux, vx patrilo do jazyka a druhé nie. Stav p by tak mal jedno slovo akceptovať a druhé nie, to DKA robiť nevie. Takže buď ich obe akceptuje alebo nie, čiže nerozoznáva daný jazyk, čím sme sa dostali do sporu s tvrdením, že jazyk je regulárny.

1. príklad

Dokážte alebo vyvráťte, že nasledujúce jazyky sú regulárne (regulárny jazyk je taký jazyk, pre ktorý vieme zostrojiť konečný automat, ktorý ho bude rozpoznávať).

(a) $L_1 = \{0^n1^n \mid n \in \mathbb{N}\}$

(b) $L_2 = \{w\#w \mid w \in \{a,b\}^*\};\ \Sigma = \{a,b,\#\}$

(c)* $L_3 = \{ww^R \mid w \in \{a,b\}^*\}$

(d) $L_4 = \{ww \mid w \in \{a,b\}^*\}$

(e) $L_5 = \{4|w|_0 = 2|w|_1 \mid w \in \{0,1\}^*\}$

(f)* $L_6 = \{a^{2^n} \mid n \in \mathbb{N}\}$

(g) $L_7 = \{a^{n^2} \mid n \in \mathbb{N}\}$

(h)* $L_8 = \{0^n1^m \mid n,m \in \mathbb{N}, n > m\}$

Riešenie príkladu 1a

Čo znamená, že je jazyk regulárny? Že vieme zostrojiť konečný stavový automat, ktorý bude tento jazyk rozpoznávať. Ak je jazyk neregulárny, takýto automat zostrojiť nevieme. Zamyslime sa, či budeme vedieť zostrojiť automat pre jazyk L. Čo by sme si museli pamätať?

Najprv by sme potrebovali mať stavy, v ktorých si pamätáme, koľko núl sme prečítali. Potom, z každého takéhoto stavu by sme išli do stavov, v ktorých by sme si počítali počet jednotiek. Je ale možné takýto automat zostrojiť pre $n \in \mathbb{N}$? Vieme povedať, aký by bol najväčší počet núl a jednotiek? Nekonečno. Náš automat má však iba konečný počet stavov, čiže tento jazyk nevie rozoznávať, a preto je neregulárny.

Dokážme to sporom, čiže dokazujeme opačné tvrdenie a to, že jazyk L je regulárny. Potom, pre takýto jazyk vieme zostrojiť deterministický automat A, ktorý má konečný počet stavov, čiže $|Q| = m$. Pre každý deterministický automat platí léma 3.12, ktorá hovorí, že ak dve slová skončia v rovnakom stave a prireťazíme k nim obom rovnaké slovo, tak obe tieto nové slová buď budú z jazyka $L(A)$, alebo nebudú z $L(A)$.

Kedy dve slová určite skončia v rovnakom stave? Ak ich je viac, ako stavov. Vyberme teda takú množinu slov, ktorá bude aspoň o jedna väčšia, ako počet stavov m. Nech sú to slová tvaru 0^k, kde $1 \leq k \leq m+1$. Vyberme z nich dve rôzne, ktoré skončia v rovnakom stave. Nevieme povedať, ktoré presne to budú, preto budeme používať premenné. Keďže nebudú rovnaké, jedno bude dlhšie ako druhé.

Nech $\hat{\delta}(q_0, 0^i) = \hat{\delta}(q_0, 0^j)$, kde $1 \leq i < j \leq m + 1$.

Potom z Lémy 3.12 vyplýva, že $0^i z \in L(A) \Leftrightarrow 0^j z \in L(A)$, pre každé $z \in \Sigma^*$ (pretože ak sme pri čítaní 0^i a 0^j skončili v rovnakom stave, keď z tohto stavu začneme čítať z, vieme skončiť iba v jednom stave, čiže $\hat{\delta}(q_0, 0^i z) = \hat{\delta}(q_0, 0^j z)$. A ak obe skončia v rovnakom stave a jedno by bolo z jazyka a druhé nie, vedeli by sme o tomto stave povedať, či je akceptačný?).

Teraz si musíme zvoliť z tak, aby jedno slovo bolo z jazyka L a druhé nie, tak získame spor s tvrdením, že jazyk je regulárny. Ako vyzerajú slová z jazyka? Majú niekoľko núl a po nich rovnaký počet jednotiek. Zvoľme preto $z = 1^i$. Potom:

- $0^i 1^i \in L(A)$
- $0^j 1^i \notin L(A)$, keďže $i < j$, čiže v tomto slove bude viac núl ako jednotiek.

Dostali sme sa do sporu s tvrdením, že jazyk L je regulárny, platí preto pôvodné tvrdenie – jazyk L je neregulárny.

Zamyslite sa, či aj jazyk $L = \{0^n 1^n \mid 0 < n < 2021\}$ je neregulárny, alebo preň vieme zostrojiť automat, ktorý ho bude rozpoznávať.

Riešenie príkladu 1b

Dokazujeme sporom. T.j. dokazujeme tvrdenie, že jazyk $L = \{w\#w \mid w \in \{a, b\}^*\};\ \Sigma = \{a, b, \#\}$ je regulárny. Potom musí existovať automat $A = \{Q, \Sigma, \delta, q_0, F\}$, pre ktorý platí že $L(A) = L$. Nech počet stav tohto (konečného) automatu je n.

Keď vezmeme slová $a, a^2, a^3, ..., a^{n+1}$, tak potom[1] musia aspoň dve slová skončiť v rovnakom stave, čiže $\hat{\delta}(q_0, x) = \hat{\delta}(q_0, y)$.

Nech $x = a^i$ a $y = a^j$, pričom $1 \leq i < j \leq n + 1$. (Sú to teda nejaké dve rôzne slová z postupnosti vyššie[2].) Čiže $\hat{\delta}(q_0, a^i) = \hat{\delta}(q_0, a^j)$. A z Lémy

[1] z Pigeonhole principle

[2] Častou chybou je, že vyberiete nejaké konkrétne slová. To spraviť nemôžeme, lebo nevieme, ktoré z nich skončia v rovnakom stave. Preto ich musíme zapísať takto všeobecnejšie a dávať pritom pozor, aby sme zaručili, že tie slová budú rôzne a z množiny väčšej ako je počet stavov automatu.

3.12 vieme, že potom platí $a^j z \in L(A) \Leftrightarrow a^i z \in L(A)$ (pretože $\hat{\delta}(q_0, a^j z) = \hat{\delta}(q_0, a^i z)$).

Zvoľme si $z = \#a^i$. Takže $\hat{\delta}(q_0, a^j\#a^i) = \hat{\delta}(q_0, a^i\#a^i)$.

- $a^j\#a^i \notin L(A)$, pretože $i < j$
- $a^i\#a^i \in L(A)$

Dostali sme sa do sporu. Preto platí pôvodné tvrdenie a jazyk L nie je regulárny.

Mohli by ste sa pýtať, prečo sme zvolili a^i, a^j a nikde v slove sa nenachádza b. Je na nás, akú množinu slov si zvolíme, nemusíme použiť všetky znaky abecedy, ale na konci potrebujeme dostať spor – čiže ak k našim slovám prireťazíme nejaké iné slovo, jedno bude z pôvodného jazyka a druhé nie. Čo okrem $\#a^i$ by sme mohli k slovám ešte prireťaziť, aby sme sa dostali do sporu? Napríklad $b^3\#a^i b^3$, $ba\#a^i ba$... Čiže to, čo dáme pred mriežku, zopakujeme aj za ňou.

Čo by sa stalo, ak by sme prireťazili $\#a^j$? Dostali by sme spor? Áno, keďže slovo $a^j\#a^j \in L(A)$, ale $a^i\#a^j \notin L(A)$.

Riešenie príkladu 1d

Jazyk nie je regulárny. Dokazujeme sporom. T.j. dokazujeme tvrdenie, že jazyk $L = \{ww \mid w \in \{a,b\}^*\}$ je regulárny. Potom musí existovať automat $A = \{Q, \Sigma, \delta, q_0, F\}$, pre ktorý platí že $L(A) = L$. Nech počet stav tohto (konečného) automatu je n.

Keď vezmeme slová $a, a^2, a^3, ..., a^{n+1}$, tak potom musia aspoň dve slová skončiť v rovnakom stave, čiže $\exists x, y \in \Sigma^*, \hat{\delta}(q_0, x) = \hat{\delta}(q_0, y)$.

Nech $x = a^i$ a $y = a^j$, pričom $1 < i < j \leq n+1$. Čiže z Lémy 3.12 $\hat{\delta}(q_0, a^i) = \hat{\delta}(q_0, a^j) \Longrightarrow \hat{\delta}(q_0, a^j z) = \hat{\delta}(q_0, a^i z)$, pre každé $z \in \Sigma^*$

Ako si zvoliť z? Možno vám intuitívne napadne, že veď stačí zvoliť a^i alebo a^j. To ale nie je správne. Prečo? Aké slová tvaru a^k budeme vedieť zapísať ako ww? Jednoducho slovo rozdelíme na polovice, teda $ww = a^{k/2}a^{k/2}$, a to vieme spraviť iba vtedy, ak k bude párne. Čo vieme povedať o dĺžke slov $a^i a^j$ a $a^j a^j$? Dĺžka prvého bude $i + j$, o čom nevieme povedať nič viac. Dĺžka druhého bude $i + i = 2i$, čiže párna. Vieme, že druhé slovo určite budeme môcť rozdeliť na dve rovnaké za sebou idúce časti (ww). Vieme s určitosťou povedať, že $i + j$ bude párne alebo nepárne? Nie, takže toto je nesprávne z.

Teraz by ste si mohli povedať, že veď tak upravíme slová, aby nám to vyšlo. Napríklad pri a^i a a^{i+1} by sme vedeli povedať čosi o párnosti dĺžky *druhého slova. Ale aj toto je nesprávne. Prečo? No lebo nami vybraté slová nemôžu závisieť jedno od druhého.*

Vráťme sa preto k pôvodným slovám a hľadajme iné z. Čo nám pomohlo v príklade 4b? Že sme vedeli ľahko určiť oddelenie slov pomocou mriežky #. Vieme tu použiť niečo iné, čo by slúžilo rovnako? Máme okrem a ešte niečo v abecede? No predsa b.

Zvoľme si $z = ba^ib$. Potom $\hat{\delta}(q_0, a^jba^ib) = \hat{\delta}(q_0, a^iba^ib)$.

- $a^jba^ib \notin L(A)$, pretože $a^jb \neq a^ib$
- $a^iba^ib \in L(A)$

Dospeli sme k sporu (buď akceptujeme obe slová, čiže aj slovo, ktoré nie je z jazyka, alebo neakceptujeme žiadne z týchto slov a teda ani slovo, ktoré patrí do jazyka), daný automat nerozoznáva jazyk L. Platí preto pôvodné tvrdenie. Jazyk L nie je regulárny.

Riešenie príkladu 1e

Dokazujeme sporom. T.j. dokazujeme tvrdenie, že jazyk $L = \{w \in \{0,1\}^* \mid 4|w|_0 = 2|w|_1\}$ je regulárny. Potom musí existovať automat $A = \{Q, \Sigma, \delta, q_0, F\}$, pre ktorý platí že $L(A) = L$. Nech počet stav tohto (konečného) automatu je n.

Teraz sa zamyslime, aké slová patria do jazyka L. Veľmi častou chybou je predpoklad, že $w = 0^41^2 \in L$. Ľahko ukážeme, že to nie je pravda: $4|w|_0 = 4|0^41^2|_0 = 16$ a $2|w|_1 = 2|0^41^2|_1 = 4$ a $4 \neq 16$.

Správne tam budú patriť napríklad slová 0^21^4, alebo 0^41^8. Všeobecne zapísané $0^{2i}1^{4i}$. (Toto, samozrejme, nie sú všetky slová z jazyka, výraz vieme zjednodušiť [3], nuly a jednotky vieme vymeniť, ale aj dať cez seba. Pri dokazovaní (ne)regulárnosti však stačí vybrať nejakú podmnožinu jazyka, ktorá je aspoň o jedna väčšia ako počet stavov automatu. Lebo ak pre nejakú podmnožinu jazyka L ukážete, že sa nedá zostrojiť automat, ktorý by rozpoznával jazyk L (pozor, jazyk, nie tú danú podmnožinu), tak ho zjavne nezostrojíme ani pre zvyšné slová.)

Ak si vyberieme slová $0^2, 0^4, ..., 0^{2n}, 0^{2(n+1)}$ bude ich viac ako stavov, preto nejaké dve slová budú musieť skončiť v rovnakom stave $\hat{\delta}(q_0, x) = \hat{\delta}(q_0, y)$.

[3] Bude správne aj 0^i1^{2i}?

Nech $x = 0^{2i}, y = 0^{2j}$, kde $1 < i < j \leq n$. Potom z Lémy 3.12 platí pre ľubovoľné z že $\hat{\delta}(q_0, 0^{2i}z) = \hat{\delta}(q_0, 0^{2j}z)$.

Zvoľme $z = 1^{4i}$. Čiže $\hat{\delta}(q_0, 0^{2i}1^{4i}) = \hat{\delta}(q_0, 0^{2j}1^{4i})$.

- $0^{2i}1^{4i} \in L(A)$, lebo $4 \,.\, 2i = 8i = 2 \,.\, 4i$
- $0^{2j}1^{4i} \notin L(A)$, lebo $4 \,.\, 2j = 8j$ a $2 \,.\, 4i = 8i$. Keďže $i < j$, potom $8j > 8i$.

Dospeli sme k sporu, platí preto pôvodné tvrdenie. Jazyk L nie je regulárny.

Riešenie príkladu 1g

Predpokladajme, že jazyk L je regulárny. Potom vieme zostrojiť automat, ktorý ho rozpoznáva, a má k stavov. Ak budeme mať (aspoň) $k+1$ slov, aspoň dve slová musia skončiť v rovnakom stave.

Majme slová tvaru a^{2m+1}, kde $m \in N$. Vidíme, že počet takýchto slov je nekonečný, preto ich určite bude viac ako k. Nech $1 \leq i < j \leq k+1$ a $\hat{\delta}(q_0,\, a^{2i+1}) = \hat{\delta}(q_0,\, a^{2j+1})$

Potom musí platiť aj $\hat{\delta}(q_0,\, a^{2i+1}x) = \hat{\delta}(q_0,\, a^{2j+1}x)$ pre akékoľvek $x \in \Sigma^*$ (z Lémy 3.12).

Zvoľme si $x = a^{j^2}$. Musí teda platiť: $a^{2i+1}a^{j^2} \in L(A) \iff a^{2j+1}a^{j^2} \in L(A)$

Upravme $a^{2j+1}a^{j^2} = a^{j^2+2j+1} = a^{(j+1)^2}$, čo určite patrí do jazyka L

Upravme $a^{2i+1}a^{j^2} = a^{j^2+2i+1}$. Vidíme, že toto slovo má viac a ako a^{j^2}, a z podmienky, že i je menšie ako j, vieme aj, že to bude menej ako v $a^{(j+1)^2}$.

Medzi a^{j^2} a $a^{(j+1)^2}$ sa nenachádza žiadna iná druhá mocnina prirodzeného čísla, preto určite $a^{j^2+2i+1} \notin L$.

Dospeli sme tak do sporu s tvrdením, že automat rozoznáva daný jazyk. V rovnakom stave skončilo slovo z jazyka, aj slovo, ktoré do jazyka nepatrí, takže automat nerozoznáva správne jazyk L. A teda jazyk L nie je regulárny.