Úvod do TI 2010

Vzorové riešenia úlohy 6.1

Bodovanie

Každý korektný dôkaz bol ohodnotený 5-timi bodmi.

Úloha

Dokážte že jazyk

$$L = \{a^{n^2} | n \in N\}$$

nie je regulárny pomocou:

a) Lemy 3.12

b) Pumpovacej lemy

a)

Lema 1 (lema 3.12) *Nech $L \subseteq \Sigma^*$ je regulárny jazyk a $x, y \in \Sigma^*$ sú také slová nad rovnakou abecedou, že $\hat{\delta}(x, q_0) = \hat{\delta}(y, q_0)$. Potom pre ľubovoľné slovo $z \in \Sigma^*$ platí*

$$xz \in L \Leftrightarrow yz \in L$$

Dôkaz spravíme sporom. Nech L je regulárny jazyk. Potom existuje konečný automat $M = (Q, \Sigma, q_0, F, \delta), L(M) = L$ ktorý ho akceptuje. Zoberme slovo $u = a^{|Q|}$. Výpočet automatu M na slove u je dlhý $|Q| + 1$ krokov a preto sa vo výpočte musia nejaký stav zopakovať. To znamená, že existujú $i < j \leq |Q|$ také, že $\hat{\delta}(a^i, q_0) = \hat{\delta}(a^j, q_0)$. Podľa lemy 3.12 pre ľubovoľné $z \in \Sigma^*$ platí, že $a^i z \in L \Leftrightarrow a^j z \in L$. Zoberme $z = a^{|Q|^2 - i}$. Zjavne $a^i z \in L$. Podľa lemy 3.12 slovo $v = a^j z = a^j a^{|Q|^2 - i} = a^{|Q|^2 + j - i}$ patrí do jazyka L. Zoberme si jeho dĺžku. $|v| = |Q|^2 + j - i$. Z toho že $i < j$ vieme, že $|v| > |Q|^2$. Z toho, že $i, j \leq |Q|$ dostaneme že $|v| \leq |Q|^2 + |Q|$. To znamená, že $|Q|^2 < |v| \leq |Q|^2 + |Q|$. Môže byť $|v|$ štvorec? Najbližšší štvorec k $|Q^2|$ je $(|Q|+1)^2 = |Q|^2 + 2|Q| + 1$ a teda $|v|$ nemôže byť štvorec a preto nemôže patriť do jazyka L. To je ale spor. Preto L nie je regulárny.

Poznámka: Nájsť také slová, na ktorých automat skončí v tom istom stave sa dá aj ináč. Napríklad si môžeme zobrať ľubovoľných $|F| + 1$ slov z jazyka L z dirichletovho princípu dostaneme, že na nejakých dvoch slovách automat M skončí v rovnakom akceptačnom stave (pozor, akceptačných stavov môže byť viac, čo ste si mnohí vo svojich riešeniach neuvedomovali).

b)

Lema 2 (Pumpovacia lema) *Nech $L \subseteq \Sigma^*$ je regulárny jazyk. Potom existuje konštanta n_0 taká, že každé slovo $w \in \Sigma^*, |w| \geq n_0$ sa dá rozdeliť na tri časti $w = yxz$ tak, že platí:*

1. $|yx| \leq n_0$
2. $|x| \geq 1$
3. $\{yx^k x | k \in N\} \subseteq L$ *alebo* $\{yx^k x | k \in N\} \cap L = \emptyset$.

Dôkaz spravíme sporom. Nech L je regulárny jazyk. Potom existuje konštanta n_0 taká, že pre každé slovo... Zoberme teda slovo $w \in L, |w| \geq n_0^2$ (poznámka – zobrali sme oveľa dlhšie slovo, ako potrebuje pumpovacia lema. Tam by stačilo slovo dĺžky n_0. Ale s týmto slovom sa nám bude lepšie dokazovať). Keďže $w \in L$, tak $w = a^{n^2}$ pre nejaké $n \geq n_0$. Podľa pumpovacej lemy ho vieme rozdeliť na tri časti $w = yxz$, také, že $|yx| \leq n_0, |x| \geq 1$. Z toho vyplýva, že $|x| \leq n_0 \leq n$. Keďže $w \in L$, tak z pumpovacej lemy dostaneme že $v = yx^2z \in L$. Ale $|v| = |w| + |x| = n^2 + |x|$. Keďže $|x| \leq n$, tak aj $|x| < 2n + 1$ a preto $|v|$ leží medzi dvoma štvorcami, $|v|$ nie je štvorec. Dostali sme sa do sporu s predpokladom a preto L nie je regulárny jazyk.

Poznámka: Niektorí z vás ste v riešení uviedli, že z pumpovacej lemy vyplýva že $n_0 = n^2$. Toto nie je pravda. Pumpovacia lema nehovorí nič o tom, aké veľké je n_0. Hovorí iba že pre každý regulárny jazyk také n_0 existuje. Z dôkazy pumpovacej lemy vyplýva, že n_0 stačí aby bolo $|Q|$ (ale pre konkrétne jazyky to môže platiť aj pre menšie čísla, ale číslo $|Q|$ je dostatočné).

Vzorové riešenia úlohy 6.2

a)

Nech M je konečný automat s n stavmi, ktorý akceptuje L. Nech $u = wxw^R$ a $|w| > n$. Potom aspoň raz nastane situácia, že sa automat M pri čítaní slova w ocitne aspoň dvakrát v rovnakom stave. Nech sa tak stane po prečítaní prvých i respektíve j znakov slova w. Bez ujmy na všeobecnosti môžeme predpokladať, že $i \neq j$. Označme si w_i prvých i znakov slova w(obdobne w_j). Teda $(q_0, w_i) \vdash_M^* = (p, \lambda)$ a $(q_0, w_j) \vdash_M^* = (p, \lambda)$. Nech v je slovo w bez prefixu w_i. Potom podľa lemy 3.12 automat M musí akceptovať aj slovo w_jvxw^R (pretože akceptuje $w_ivxw^R = wxw^R$), ktoré nepatrí do jazyka L. To je spor, teda jazyk L nie je regulárny.

b)

Tu si treba uvedomiť, že $\{\lambda x \lambda^R | x \in \{0,1\}^*\} \subseteq \{wxw^R | x, w \in \{0,1\}^*\}$. Teda $\{0,1\}^* \subseteq \{wxw^R | x, w \in \{0,1\}^*\}$. Teda aj $\{0,1\}^* = \{wxw^R | x, w \in \{0,1\}^*\} = L$. Takýto jazyk je regulárny a akceptuje ho ľubovoľný konečný automat, ktorý má všetky stavy akceptačné.

Časté chyby: Treba dodržiavať konzistentnosť označenia. Keď si raz označíte $v = wxw^R$, tak nemôže zároveň platiť $v = wx$. Vo viacerých riešeniach ste postupovali nasledovne. Najskôr ste zobrali slovo $v_1 = w_1x_1w_1^R$, kde $w_1 = 10$ a $x = 00$. Slovo v_1 patrí do jazyka L. Potom ste zobrali slovo $v_2 = w_2x_2w_2^R$, kde $w_2 = 1$ a $x = 0000$. Na základe toho ste tvrdili, že v_2 nepatrí do jazyka L. Ale $v_1 = v_2$, tak ako môže to isté slovo patriť aj nepatriť do toho istého jazyka?

Vzorové riešenie úlohy 6.3

Hľadáme (deterministický) automat s najmenším možným počnom stavov, ktorý akceptuje jazyk obsahujúci práve 1000 slov nad $\{a, b\}$. Naša úloha sa dá rozdeliť na dve časti: najprv zostrojíme automat s 11 stavmi (horný odhad – ukážeme že výsledok je najviac 11) a potom dokážeme, že na menej ako 11 stavov to nejde (dolný odhad).

Horný odhad Ukážeme, že automat $M = (\{q_0, \dots, q_{10}\}, \{a, b\}, \delta, q_0, \{q_3, q_5, q_6, q_7, q_8, q_9\})$, kde

$$
\begin{aligned}
\delta(q_0, c) &= q_1, \quad c \in \{a, b\},\\
\delta(q_1, c) &= q_2, \quad c \in \{a, b\},\\
\delta(q_2, c) &= q_3, \quad c \in \{a, b\},\\
\delta(q_3, c) &= q_4, \quad c \in \{a, b\},\\
\delta(q_4, c) &= q_5, \quad c \in \{a, b\},\\
\delta(q_5, c) &= q_6, \quad c \in \{a, b\},\\
\delta(q_6, c) &= q_7, \quad c \in \{a, b\},\\
\delta(q_7, c) &= q_8, \quad c \in \{a, b\},\\
\delta(q_8, c) &= q_8, \quad c \in \{a, b\},\\
\delta(q_9, c) &= q_{10}, \quad c \in \{a, b\},\\
\delta(q_{10}, c) &= q_{10}, \quad c \in \{a, b\},
\end{aligned}
$$

akceptuje jazyk, ktorý má práve 1000 slov. Automat má jediný cyklus, v stave q_10. Ukážeme si, ako vyzerajú triedy $KL[q_i]$: Trieda $KL[q_0]$ obsahuje zjavne len prázdne slovo. Do stavu q_1 sa dá dostať len zo stavu q_0, na písmená a a b. Preto $KL[q_1] = KL[q_0] \cdot \{a, b\} = \{a, b\}$. Podobným spôsobom môžeme popísať všetky triedy $KL[q_i]$ na základe $KL[q_{i-1}]$ a dostávame $KL[q_i] = \{a, b\}^i$. Rozdiel bude až v stave q_{10}, ktorý nás však už nezaujíma, pretože nie je akceptačný. (A ani sa z neho nevieme dostať do žiadneho akceptačného stavu.) Vidíme, že $|KL[q_i]| = 2^i$ a preto $|L(M)| = |KL[q_3]| + |KL[q_5]| + |KL[q_6]| + |KL[q_7]| + |KL[q_8]| + |KL[q_9]| = 2^3 + 2^5 + 2^6 + 2^7 + 2^8 + 2^9 = 1000$. Iným spôsobom ako ukázať, že $L(M) = \{w | |w| \in \{3, 5, 6, 7, 8, 9\}\}$ je ukázať obe inklúzie. Myšlienkou jednej inklúzie je, že každé slovo správnej dĺžky skončí v akceptačnom stave. Myšlienkou opačnej inkúzie je, že každé akceptované slovo má správnu dĺžku. Zamyslite sa nad tým, ako by vyzeral poriadny (formálny) dôkaz oboch inklúzií.

Dolný odhad Chceme akceptovať konečný jazyk, preto musí platiť, že medzi akceptačnými stavmi a stavmi, z ktorých sa dá dostať do akceptačných stavov nie je cyklus. Cyklus môže byť len v stavoch, z ktorých sa už nevieme dostať do akceptačného stavu. Navyše takýto „odpadový“ stav musí existovať, pretože musíme vedieť dočítať každý vstup. Predpokladajme že máme menej ako 11 stavov, jeden z nich je odpadový, preto máme najviac 9 neodpadových stavov. Na týchto stavoch nie je cyklus, preto môžeme akceptovať slová dĺžky maximálne 8. (Zamyslite sa nad tým prečo. Pomôcka: v minulej sade domácich úloh bol príklad, ktorý sa venoval minimálnemu počtu stavov, ak chceme akceptovať slová istej dĺžky.) Koľko je slov, ktoré majú dĺžku najviac 8? Je ich $2^0 + 2^1 + \cdots + 2^8 = 511$. Z toho vyplýva, že 10 stavový automat nemôže akceptovať konečný jazyk obsahujúci 1000 slov.

Hodnotenie, poznámky Väčšina z vás nerobila dolný odhad, čo vám nejaké bodíky ubralo. Niektorí ani neukazovali, že automat ktorý zostojili akceptuje práve 1000 slov, čo je tiež chyba. Drvivej väčšine sa však podarilo aspoň nájsť správny automat.