Richard Kráľovič

Šiesta sada domácich úloh

Úloha č. 6.7:

Jazyk $L = \{a^z | z \text{ je zložené číslo}\}$ nie je regulárny.

Lema 1: Prvočísel je nekonečne veľa.

Dôkaz: Nech by bolo prvočísel konečne veľa (označme ich počet n). Potom ich môžme označiť $p_1, p_2 \dots p_n$. Zjavne číslo $1 + \prod_{i=1}^{n} p_i$ je väčšie ako ľubovoľné p_i, podľa predpokladu teda nie je prvočíslo. Toto číslo zjavne dáva zvyšok 1 po delení p_i pre všetky $1 \leq i \leq n$. Nie je teda deliteľné žiadnym prvočíslom. Je teda deliteľné iba číslom 1 a sebou samým, čo je však spor s tým, že nie je prvočíslo. Prvočísel je teda nekonečne veľa.

Lema 2: Pre všetky prirodzené n existuje interval n prirodzených čísel, medzi ktorými sa nenachádza žiadne prvočíslo.

Dôkaz: Uvažujme interval prirodzených čísel $\langle (n+1)! + 2, (n+1)! + n + 1 \rangle$. Zjavne tento interval obsahuje n prirodzených čísel. Nech p patrí do tohoto intervalu, potom $p = (n+1)! + i$, kde $2 \leq i \leq n+1$. Potom

$$p = i + \prod_{j=1}^{n+1} j = i\left(1 + \left(\prod_{j=1}^{i-1} j\right)\left(\prod_{j=i+1}^{n+1} j\right)\right)$$

Čo je pre $n \geq 1$ pre každé i zložené číslo. Dokazovaná lema teda platí.

Dôsledok: Z lemy 1 a lemy 2 vyplýva, že pre ľubovoľné číslo n existuje prirodznené číslo a také, že pre všetky i také, že $a \leq i < a + n$, je i zložené číslo a číslo $a + n$ je prvočíslo.

Teraz sporom dokážeme, že jazyk L nie je regulárny. Nech by jazyk L bol regulárny. Potom existuje n-stavový deterministický konečný automat A, ktorý tento jazyk rozpoznáva. Podľa dôsledku lemy 1 a lemy 2 existuje prirodzené číslo k také, že pre všetky i také, že $k \leq i < k+n$, je i zložené číslo a číslo $k+n$ je prvočíslo. Potom automat nesmie akceptovať slovo a^{k+n}. Pretože automat je deterministický, pre vstupné slovo a^{k+n} existuje práve jeden výpočet, ktorý môžme zapísať nasledovne: $\left(q_0, a^{k+n}\right) \vdash_A \left(q_1, a^{k+n-1}\right) \vdash_A \cdots \vdash_A (q_k, a^n) \vdash_A \cdots \vdash_A (q_{k+n}, \varepsilon)$, pričom q_0 je počiatočný stav automatu a q_{k+n} nie je akceptačný stav automatu.

Pretože automat A je n-stavový, z Dirichletovho princípu vyplýva, že existujú prirodzené čísla i, j také, že $k \leq i < j \leq k+n$ a $q_i = q_j$. Výpočet automatu potom môžme zapísať nasledovne: $\left(q_0, a^{k+n}\right) \vdash_A \left(q_1, a^{k+n-1}\right) \vdash_A \cdots \vdash_A (q_k, a^n) \vdash_A \cdots \vdash_A \left(q_i, a^{k+n-i}\right) \vdash_A \cdots \vdash_A \left(q_j, a^{k+n-j}\right) \vdash_A \cdots \vdash_A (q_{k+n}, \varepsilon)$. Potom ale výpočet automatu pre slovo $a^{k+n-(j-i)}$ je nasledovný: $\left(q_0, a^{k+n-(j-i)}\right) \vdash_A \left(q_1, a^{k+n-1-(j-i)}\right) \vdash_A \cdots \vdash_A \left(q_k, a^{n-(j-i)}\right) \vdash_A \cdots \vdash_A \left(q_i, a^{k+n-(j-i)}\right) = \left(q_j, a^{k+n-j}\right) \vdash_A \cdots \vdash_A (q_{k+n}, \varepsilon)$, automat A teda neakceptuje ani slovo $a^{k+n-(j-i)}$.

Pretože platí $j \leq k+n$, $k \leq i$ a $i < j$, platí aj $0 < j - i \leq n$, teda aj $k + n > k + n - (j - i) \geq k$. Podľa predpokladu ale všetky čísla spĺňajúce túto nerovnosť sú zložené, automat A teda neakceptuje slovo a^x, kde x je nejaké zložené číslo. To je spor s predpokladom, že automat A akceptuje všetky slová tvaru a^x, kde x je zložené číslo.

Dokázali sme, teda, že jazyk L nie je rozpoznávaný nijakým konečným automatom, nie je teda regulárny.

Úloha č. 6.11:

Bezkontextová gramatika generujúca jazyk $L = \{w | w \in \{a,b\}^*, \#_a w = \#_b w\}$ je napr. táto: $G = (\{\sigma\}, \{a,b\}, P, \sigma)$, kde množina pravidiel P je nasledovná:

1. $\sigma \to \varepsilon$
2. $\sigma \to \sigma\sigma$
3. $\sigma \to a\sigma b$
4. $\sigma \to b\sigma a$

Najskôr ukážeme, že všetky slová, ktoré gramatika vygeneruje, patria do L. Budeme dokazovať silnejšie tvrdenie, že pre každú vetnú formu w gramatiky G, platí $\#_a w = \#_b w$. Dôkaz vykonáme indukciou vzhľadom na počet krokov odvodenia.

1°: Vetná forma odvodená na 0 krokov odvodenia je σ, pre ktorú zjavne platí $\#_a\sigma = 0 = \#_b\sigma$.

2°: Nech pre všetky vetné formy vygenerované na $n-1$ krokov odvodenia platí dokazované tvrdenie. Z definície pravidiel gramatiky G vyplýva, že ak w je vetná forma odvodená na n krokov odvodenia, existujú $u, v, z \in \{a, b, \sigma\}^*$, kde z je vetná forma G odvodená na $n-1$ krokov odvodenia také, že $z = u\sigma v$ a $w = uv$, $w = u\sigma\sigma v$, $w = ua\sigma bv$ resp. $w = ub\sigma av$. Zjavne platí $\#_a z = \#_a u + \#_a v$ a $\#_b z = \#_b u + \#_b v$. Podľa indukčného predpokladu teda platí $\#_a u + \#_a v = \#_b u + \#_b v$. Zjavne platí, že $\#_a w = \#_a u + \#_a v$, $\#_a w = \#_a u + \#_a v$, $\#_a w = \#_a u + \#_a v + 1$ resp. $\#_a w = \#_a u + \#_a v + 1$. Taktiež platí $\#_b w = \#_b u + \#_b v$, $\#_b w = \#_b u + \#_b v$, $\#_b w = \#_b u + \#_b v + 1$ resp. $\#_b w = \#_b u + \#_b v + 1$. Podľa indukčného predpokladu vo všetkých prípadoch platí $\#_a w = \#_b w$, teda aj pre vetnú formu w platí dokazované tvrdenie.

Týmto sme dokázali, ze všetky slová generované gramatikou G patria do jazyka L. Teraz dokážeme, že všetky slova z jazyka L môže gramatika vygenerovať.

Lema 1: Nech pre slovo w platí $\#_a w = \#_b w$ a nech je slovo w tvaru axa, kde $x \in \{a, b\}^*$. Potom existujú $u, v \in \{a, b\}^+$ také, že $w = uv$ a zároveň $\#_a u = \#_b u$ (Z toho vyplýva, že aj $\#_a v = \#_b v$).

Dôkaz: Nech x je prefix slova w a nech $|x| = 1$. Potom zjavne $\#_a x - \#_b x = 1$. Nech y je prefix slova w a nech $|y| = |w| - 1$. Potom zjavne $\#_a y - \#_b y = -1$, lebo $\#_a w = \#_b w$. Ak q_1 a q_2 sú prefixy slova w také, že $|q_2| = |q_1| + 1$, potom zjavne $\#_a q_1 - \#_b q_1 = \#_a q_2 - \#_b q_2 \pm 1$. Ak by neexistovali $u, v \in \{a, b\}^+$ také, že $w = uv$ a zároveň $\#_a u - \#_b u = 0$, muselo by pre všetky prefixy q slova w, kde $1 \leq |q| < |w|$ platiť $\#_a q - \#_b q > 0$, teda aj $\#_a y - \#_b y > 0$, čo je spor. Musia teda existovať také $u, v \in \{a, b\}^+$ také, že $w = uv$ a zároveň $\#_a u = \#_b u$.

Lema 2: Nech pre slovo w platí $\#_a w = \#_b w$ a nech je slovo w tvaru bxb, kde $x \in \{a, b\}^*$. Potom existujú $u, v \in \{a, b\}^+$ také, že $w = uv$ a zároveň $\#_a u = \#_b u$ (Z toho vyplýva, že aj $\#_a v = \#_b v$).

Dôkaz tejto lemy je analogický ako dôkaz lemy 1.

Teraz indukciou vzhľadom na dĺžku slova dokážeme, že všetky slová w patriace do jazyka L môže gramatika vygenerovať z počiatočného neterminálu.

1°: Nech slovo w obsahuje 0 písmen a aj b. Potom $w = \varepsilon$ a môže byť vygenerované z σ pomocou pravidla č. 1.

2°: Predpokladajme, že všetky slová obsahujúce najviac $n-1$ písmen a a rovnako veľa písmen b môžu byť vygenerované z počiatočného neterminalu σ. Potom pre slovo w obsahujúce n písmen a aj b platí jedna z nasledovných možností:

   1. $w = aub$. Potom u obsahuje $n-1$ písmen a aj b. Z počiatočného neterminálu môže G odvodiť pravidlom č. 3 vetnú formu $a\sigma b$. Podľa indukčného predpokladu môže zo σ odvodiť u, môže teda odvodiť aj aub.

   2. $w = bua$. Potom u obsahuje $n-1$ písmen a aj b. Z počiatočného neterminálu môže G odvodiť pravidlom č. 4 vetnú formu $b\sigma a$. Podľa indukčného predpokladu môže zo σ odvodiť u, môže teda odvodiť aj bua.

   3. $w = aua$ resp $w = bub$. Potom podľa lemy 1 resp. lemy 2 existujú slová $q, v \neq \varepsilon$ také, že $w = qv$ a zároveň $\#_a q = \#_b q$ a $\#_a v = \#_b v$. Pretože $q, v \neq \varepsilon$, $|q| < |w|$

aj $|v| < |w|$, teda podľa indukčného predpokladu zo σ môže gramatika G vygenerovať v aj q. Zo σ môže gramatika G vygenerovať pravidlom č. 2 vetnú formu $\sigma\sigma$, môže teda vygenerovať aj slovo $qv = w$

Dokázali sme, že každé slovo vygenerované gramatikou G patrí do L a zároveň každé slovo z L môže G vygenerovať, teda $L(G) = L$.