martin homola, 2i2

1. sada domácich úloh - úloha 8
Úlohou je zostrojiť kontextovú gramatiku pre jazyk

$$L = \{w \mid w \in \{a,b,c\}^* \text{ také, že } \#_a w = \#_b w = \#_c w\}$$

Riešenie:
Gramatiku $G = (N, T, P, \sigma)$ definujeme nasledovne:

$$N = \{\sigma, \alpha, \beta, \gamma\}$$

$$T = \{a, b, c\}$$

$$\begin{array}{rlll} P = \{ & \sigma \rightarrow \sigma\sigma & (1) & \\ & \sigma \rightarrow \alpha & (2) & \\ & \sigma \rightarrow \varepsilon & (3) & \\ & \alpha \rightarrow a\beta & (4) & \\ & \beta \rightarrow b\gamma & (5) & \\ & \gamma \rightarrow c & (6) & \} \cup \end{array}$$

$$\cup\ P_1 = \{\ x\xi \rightarrow \xi x \mid x \in T,\ \xi \in N - \{\sigma\}\ \} \cup$$

$$\cup\ P_2 = \{\ \xi x \rightarrow x\xi \mid x \in T,\ \xi \in N - \{\sigma\}\ \}$$

Jazyk $L \in L_{ECS}$, teda aj G je rozšírená kontextová gramatika. Kontextovú gramatiku pre jazyk $L - \{\varepsilon\} \in L_{CS}$ získame vylúčením pravidla $\sigma \rightarrow \varepsilon$ z množiny pravidiel P.

Tvrdím, že $L(G) = L$.

Dôkaz:

A. $L \supseteq L(G)$**:**
Matematickou indukciou vzhľadom na dĺžku slova $|w| = n$.
1° $w \in L(G),\ |w| = n = 0 \ \Rightarrow\ w = \varepsilon \in L$
2° Pre všetky slová jazyku L platí: $3 \ \Big|\ |w|$. Indukčný krok zohľadňuje túto skutočnosť, z indukčného predpokladu pre dĺku slova k vyplýva dôsledok pre dĺžku slova $k+3$.
Predpokladajme teda, že pre ľubovoľné slovo $w \in L(G)$, $|w| = k$ platí $w \in L$.
Nech teraz $w \in L(G)$, $|w| = k+3$. Potom existuje pravé krajné odvodenie slova w v gramatike G:

$$o_1 :\ \sigma \Rightarrow^* w.$$

Niekoľko posledných krokov odvodenia o_1 zrejme vyzerá nasledovne:

$$o_1 :\ \sigma \Rightarrow^* \sigma^i u\gamma v \Rightarrow \sigma^i ucv \Rightarrow^* ucv = w;\ u, v \in T^*, i \in \{0, 1, ...\}$$

T.j. niektoré c vznikne ako posledný terminálny symbol z neterminálu γ použitím pravidla (6). Zvyšné neterminály su všetky σ (o_1 je pravé krajné odvodenie). Postupným aplikovanímpravidiel (1) a (3) sa všetky zmenia na ε.

Ďalej neterminál γ vznikol reťazcom krokov z neterminálu σ, niekoľko ďalších predchádzajúcich kokov zrejme vyzerá nasledovne:

$$o_1: \ \sigma \Rightarrow^* \boldsymbol{\sigma^i \sigma z_1} \overset{\mathrm{k}}{\boldsymbol{\Rightarrow}} \boldsymbol{\sigma^i \alpha z_1} \overset{\mathrm{l}}{\Rightarrow^*} \sigma^i z_2 \alpha z_3 \Rightarrow \sigma^i z_2 a \beta z_3 \overset{\mathrm{m}}{\Rightarrow^*} \sigma^i z_4 \beta z_5 \Rightarrow \sigma^i z_4 b \gamma z_5 \overset{\mathrm{n}}{\Rightarrow^*} \sigma^i u \gamma v \Rightarrow$$
$$\Rightarrow \sigma^i ucv \Rightarrow^* ucv = w; \ z_j, u, v \in T^*; \ i \in \{0,1,...\}; \ z_1 = z_2 z_3; \ z_2 a z_3 = z_4 z_5; \ z_4 b z_5 = uv.$$

Kde v kroku l sa niekoľko-krát použijú pravidlá tvaru $x\alpha \to \alpha x \in P_1, x \in T$ a $\alpha x \to x\alpha \in P_2, x \in T$. V krokoch m, n analogicky pre neterminály β a γ. Inými slovami, v krokoch l, m a n nepribudnú do spracovávanej vetnej formy žiadne terminály.
Pozrime sa bližšie na slovo $z_1 \in T^*$. Pravé krajné odvodenie o_2 tohto slova v gramatike G ľahko získame z odvodenia o_1 tak, že v kroku k namiesto pravidla (2) použijeme $i+1$-krát pravidlo (3). Dostaneme:

$$o_2: \ \sigma \Rightarrow^* \sigma^i \sigma z_1 \Rightarrow^{i+1} z_1$$

Z o_1 vyplýva, že $w = y_1 a_1 y_2 a_2 y_3 a_3 y_4$, kde $y_i \in T^*$, $a_j \in T$ a platí $z_1 = y_1 y_2 y_3 y_4$ a (a_1, a_2, a_3) je permutáciou (a, b, c). Teda $z_1 \in L(G), |z_1| = k \overset{\mathrm{IP}}{\Rightarrow} z_1 \in L$. Ďalej slovo w má práve o 1 symbol a viac ako slovo z_1, to isté platí aj pre symboly b a c. Potom $|w| \in L$.

B. $L \subseteq L(G)$**:**
Matematickou indukciou vzhľadom na dĺžku slova $|w| = n$.
1° $w \in L, \ |w| = n = 0 \ \Rightarrow \ w = \varepsilon \in L(G)$
2° Podobne ako v prvej časi dôkazu, indukčný krok zohľadňuje skutočnosť: $(\forall w \in L) \ 3 \ \Big| \ |w|$.
Predpokladajme teda, že pre každé slovo $w \in L$, $|w| = k$ platí $w \in L(G)$.
Nech teraz $w \in L$, $|w| = k+3$. Nech prvé a zľava v slove w je na l-tej pozícii, prvé b je na m-tej a prvé c je na n-tej pozícii. Predpokladajme, že $l < m < n$ (v ostatných prípadoch je postup analogický).
Vymažme teraz v slove w prvé a, b a c zľava. Získame tak nové slovo v, $|v| = k$ a (keďže sme zmazali rovnaký počet z každého druhu symbolov) $v \in L$. Teda existuje pravé krajné odvodenie slova v v gramatike G:

$$o_3: \ \sigma \Rightarrow^* v.$$

Podobne ako v časti A., ľahko demaskujeme niekoľko posledných krokov odvodenia o_3:

$$o_3: \ \sigma \Rightarrow^* \boldsymbol{\sigma^i \sigma u} \overset{\mathrm{k}}{\boldsymbol{\Rightarrow}} \boldsymbol{\sigma^i \alpha u} \Rightarrow^* \sigma^i z_1 \alpha z_2 \Rightarrow \sigma^i z_1 a \beta z_2 \Rightarrow^* \sigma^i z_3 \beta z_4 \Rightarrow \sigma^i z_3 b \gamma z_4 \Rightarrow^* \sigma^i z_5 \gamma z_6 \Rightarrow$$
$$\Rightarrow \sigma^i z_5 c z_6 \Rightarrow^* z_5 c z_6 = v; \ z_j, u \in T^*; \ i \in \{0,1,...\}; \ u = z_1 z_2; \ z_1 a z_2 = z_3 z_4; \ z_3 b z_4 = z_5 z_6.$$

Odvodenie o_3 možno v kroku k modifikovať tak, že miesto pravidla (2) použijeme pravidlo (1) a nasledovne pravidlo (2). Takto získame odvodenie o_4:

$$o_4: \ \sigma \Rightarrow^* \boldsymbol{\sigma^i \sigma u} \overset{\mathrm{k}}{\boldsymbol{\Rightarrow}} \boldsymbol{\sigma^i \sigma \sigma u \Rightarrow \sigma^i \sigma \alpha u} \Rightarrow^* \sigma^i \sigma v; \ u \in T^*; \ i \in \{1,2,...\}.$$

Pokračujme v odvodení o_4:

$$o_4: \ \sigma \Rightarrow^* \sigma^i \sigma v \Rightarrow \underbrace{\sigma^i \alpha v \Rightarrow^{l-1} \sigma^i y_1 \alpha y_2}_{\alpha x \to x\alpha \ \in \ P_2; \ x \in T} \Rightarrow \underbrace{\sigma^i y_1 a \beta y_2 \Rightarrow^{m-l-1} \sigma^i y_3 \beta y_4}_{\beta x \to x\beta \ \in \ P_2; \ x \in T} \Rightarrow$$
$$\Rightarrow \underbrace{\sigma^i y_3 b \gamma y_4 \Rightarrow^{n-m-l-1} \sigma^i y_5 \gamma y_6}_{\gamma x \to x\gamma \ \in \ P_2; \ x \in T} \Rightarrow \sigma^i y_5 c y_6 = \sigma^i w \Rightarrow^i w;$$
$$y_j \in T^*; \ i \in \{0,1,...\}; \ v = y_1 y_2; \ y_1 a y_2 = y_3 y_4; \ y_3 b y_4 = y_5 y_6$$

Teda existuje odvodenie slova w v gramatike G, t.j. $w \in L(G)$.

Záver: Dokázali sme, že platí $L \in L(G)$ a zároveň $L(G) \in L$, teda $L(G) = L$.