IB102 – úkol 8 – řešení **Odevzdání: 28. 11. 2011**

Vypracoval: James Bond UČO: 007
Skupina: MI6

1. [**2 body**] Mějme následující jazyk nad abecedou $\{a, b, c, d\}$:

$$L = \{a^n b^m c^r d^s \mid n + m = r + s \text{ nebo } n = s\}$$

Sestrojte jednoznačnou bezkontextovou gramatiku generující tento jazyk. Stručně zdůvodněte, proč je Vaše gramatika jednoznačná.

(Pokud nevíte, jak sestrojit jednoznačnou gramatiku, zkuste sestrojit alespoň nějakou bezkontextovou gramatiku generující tento jazyk. V tom případě bude Vaše řešení hodnoceno maximálně 1 bodem.)

Řešení: Hledaná gramatika může vypadat například takto:
$G = (\{S, B, C, X, Y, Z\}, \{a, b, c, d\}, P, S)$, kde

$$P = \{ \begin{array}{rcl} S & \to & aSd \mid B \mid C \mid BC \mid bXd \mid aYc \mid \varepsilon, \\ B & \to & b \mid bB, \\ C & \to & c \mid cC, \\ X & \to & bXd \mid Z, \\ Y & \to & aYc \mid Z, \\ Z & \to & bZc \mid \varepsilon \end{array} \}.$$

Zdůvodnění jednoznačnosti této gramatiky je následující:

Nejprve se podíváme na slova tvaru $a^n b^m c^r d^s$, kde $n = s$. Zřejmě se při odvození takovýchto slov nikdy nepoužije pravidlo $S \to bXd$ ani pravidlo $S \to aYc$, protože po jejich použití už nikdy není počet a a d ve větné formě stejný. Slovo tvaru $a^n b^m c^r d^s$, kde $n = s$, tedy musí vzniknout právě jedním z těchto odvození:

- $S \Rightarrow^n a^n S d^n \Rightarrow a^n BC d^n \Rightarrow^* a^n b^m c^r d^n$ (pokud $m > 0$ a $r > 0$)
- $S \Rightarrow^n a^n S d^n \Rightarrow a^n B d^n \Rightarrow^* a^n b^m d^n$ (pokud $m > 0$ a $r = 0$)
- $S \Rightarrow^n a^n S d^n \Rightarrow a^n C d^n \Rightarrow^* a^n c^r d^n$ (pokud $m = 0$ a $r > 0$)
- $S \Rightarrow^n a^n S d^n \Rightarrow a^n d^n$ (pokud $m = r = 0$)

Protože odvození $B \Rightarrow b^m$ $(m > 0)$ i odvození $C \Rightarrow c^r$ $(r > 0)$ je vždy jenom jedno (pro dané m nebo r), znamená to, že pro každé slovo tvaru $a^n b^m c^r d^s$, kde $n = s$, existuje právě jeden derivační strom.

Odvození slov tvaru $a^n b^m c^r d^s$, kde $n < s$ a $n + m = r + s$, musí nutně vypadat takto:

$$S \Rightarrow^n a^n S d^n \Rightarrow a^n bXd^{n+1} \Rightarrow^* a^n b^{s-n} X d^s \Rightarrow a^n b^{s-n} Z d^s \Rightarrow^* a^n b^{s-n+r} Z c^r d^s \Rightarrow a^n b^{s-n+r} c^r d^s$$

a odvození slov tvaru $a^n b^m c^r d^s$, kde $n > s$ a $n + m = r + s$, musí nutně vypadat takto:

$$S \Rightarrow^n a^s S d^s \Rightarrow a^{s+1} Y c d^s \Rightarrow^* a^n Y c^{n-s} d^s \Rightarrow a^n Z c^{n-s} d^s \Rightarrow^* a^n b^m Z c^{n-s+m} d^s \Rightarrow a^n b^m c^{n-s+m} d^s$$

V obou případech jde zřejmě o jedno jediné možné odvození, pro každé slovo tvaru $a^n b^m c^r d^s$, kde $n \neq s$ a $n + m = r + s$, existuje tedy právě jeden derivační strom.

IB102 – úkol 8 – řešení **Odevzdání: 28. 11. 2011**

Vypracoval: James Bond UČO: 007
Skupina: MI6

2. [**2 body**] Mějme gramatiku $G = (\{S, A, B, C, D, E, F\}, \{a, b, c\}, P, S)$, kde

$$\begin{array}{rlcl} P = \{ & S & \rightarrow & A \mid CD \mid F, \\ & A & \rightarrow & BA \mid CAD, \\ & B & \rightarrow & b \mid bb, \\ & C & \rightarrow & \varepsilon \mid CD, \\ & D & \rightarrow & \varepsilon \mid D, \\ & E & \rightarrow & aE \mid bCb \mid cFc, \\ & F & \rightarrow & ab \mid aDaDaS \mid EC \quad \}. \end{array}$$

Sestrojte *vlastní* gramatiku G' *bez jednoduchých pravidel* takovou, že $L(G) = L(G')$. K jejímu sestrojení použijte algoritmů z přednášky.

Řešení: Nejprve odstraníme nenormované a nedosažitelné neterminály. Při napočítání normovaných neterminálů postupně dostáváme: $N_1 = \{B, C, D, F\}$, $N_2 = \{B, C, D, F, E, S\}$, $N_3 = N_2 = N_e$. Při počítání dosažitelných neterminálů dostáváme: $N_0 = \{S\}$, $N_1 = \{S, C, D, E, F\}$, $N_2 = N_1 = N'$. Nová gramatika tedy vypadá takto:
$G_1 = (\{S, C, D, E, F\}, \{a, b, c\}, P_1, S)$, kde

$$\begin{array}{rlcl} P_1 = \{ & S & \rightarrow & CD \mid F, \\ & C & \rightarrow & \varepsilon \mid CD, \\ & D & \rightarrow & \varepsilon \mid D, \\ & E & \rightarrow & aE \mid bCb \mid cFc, \\ & F & \rightarrow & ab \mid aDaDaS \mid EC \quad \}. \end{array}$$

Dále odstraníme ε-pravidla. Nejprve napočítáme $N_\varepsilon = \{C, D, S\}$ a potom podle algoritmu z přednášky dostaneme následující gramatiku: $G_2 = (\{S', S, C, D, E, F\}, \{a, b, c\}, P_2, S')$, kde

$$\begin{array}{rlcl} P_2 = \{ & S' & \rightarrow & S \mid \varepsilon, \\ & S & \rightarrow & CD \mid F \mid C \mid D, \\ & C & \rightarrow & CD \mid C \mid D, \\ & D & \rightarrow & D, \\ & E & \rightarrow & aE \mid bCb \mid cFc \mid bb, \\ & F & \rightarrow & ab \mid aDaDaS \mid EC \mid aaDaS \mid aDaaS \mid aDaDa \mid aaaS \mid aDaa \mid aaDa \mid aaa \mid E \quad \}. \end{array}$$

Tím nám mohly vzniknout (a taky vznikly) zbytečné neterminály, takže je před dalšími úpravami opět odstraníme. Zřejmě $N'' = \{S', E, F, S\}$ je množina normovaných a dosažitelných neterminálů G_2, Dostáváme tedy gramatiku: $G_3 = (\{S', S, E, F\}, \{a, b, c\}, P_3, S')$, kde

$$\begin{array}{rlcl} P_3 = \{ & S' & \rightarrow & S \mid \varepsilon, \\ & S & \rightarrow & F, \\ & E & \rightarrow & aE \mid cFc \mid bb, \\ & F & \rightarrow & ab \mid aaaS \mid aaa \mid E \quad \}. \end{array}$$

Dále odstraníme jednoduchá pravidla. Zřejmě $N'_S = \{S', S, F, E\}$, $N_S = \{S, F, E\}$, $N_E = \{E\}$, $N_F = \{F, E\}$. Provedením algoritmu z přednášky dostáváme gramatiku: $G_4 = (\{S', S, E, F\}, \{a, b, c\}, P_4, S')$, kde

$$
\begin{array}{rrcl}
P_4 = \{ & S' & \to & \varepsilon \mid aE \mid cFc \mid bb \mid ab \mid aaaS \mid aaa, \\
& S & \to & aE \mid cFc \mid bb \mid ab \mid aaaS \mid aaa, \\
& E & \to & aE \mid cFc \mid bb, \\
& F & \to & aE \mid cFc \mid bb \mid ab \mid aaaS \mid aaa \quad \}.
\end{array}
$$

Tím nám mohly vzniknout nedosažitelné neterminály. Snadno se ale ověří, že množina dosažitelných neterminálů gramatiky G_4 je $N''' = \{S', S, E, F\}$, a tedy výsledná gramatika je $G' = G_4$.