IB102 – úkol 10 – řešení **Odevzdání: 6. 12. 2010**

Vypracoval: James Bond UČO: 007
Skupina: MI6

1. [**2 body**] Mějme gramatiku $G = (\{S, X, Y, Z\}, \{a, b, c\}, P, S)$, kde

$$\begin{array}{rrcl} P = \{ & S & \to & XZX, \\ & X & \to & Xbc \mid Ybc, \\ & Y & \to & aa \mid Saa, \\ & Z & \to & ZZb \mid c \ \ \}. \end{array}$$

Ke gramatice G sestrojte (použitím algoritmů z přednášky) ekvivalentní gramatiku v Greibachové normální formě.

Řešení: Gramatika je zřejmě redukovaná, bez jednoduchých a ε-pravidel. Začneme tedy rovnou odstraněním levé rekurze. Zvolíme si např. následující uspořádání:

$$S < X < Y < Z$$

Dále postupujeme podle algoritmu z přednášky.

1. Neterminál S neobsahuje přímou levou rekurzi, jeho pravidla tedy ponecháme beze změny.

2. Neterminál X neobsahuje pravidla, jejichž pravá strana by začínala S, ale obsahuje přímou levou rekurzi. Tu odstraníme:

$$X \to Ybc \mid YbcX'$$
$$X' \to bc \mid bcX'$$

3. Neterminál Y obsahuje pravidlo $Y \to Saa$, to odstraníme a dostaneme:

$$Y \to aa \mid XZXaa$$

   Dále odstraníme pravidlo $Y \to XZXaa$:

$$Y \to aa \mid YbcZXaa \mid YbcX'ZXaa$$

   Nakonec odstraníme přímou levou rekurzi:

$$Y \to aa \mid aaY'$$
$$Y' \to bcZXaa \mid bcX'ZXaa \mid bcZXaaY' \mid bcX'ZXaaY'$$

4. Neterminál Z obsahuje pouze přímou levou rekurzi, tu odstraníme:

$$Z \to c \mid cZ'$$
$$Z' \to Zb \mid ZbZ'$$

Jako výsledek odstranění levé rekurze dostáváme gramatiku s těmito pravidly:

$$\begin{array}{rcl}
S & \rightarrow & XZX \\
X & \rightarrow & Ybc \mid YbcX' \\
X' & \rightarrow & bc \mid bcX' \\
Y & \rightarrow & aa \mid aaY' \\
Y' & \rightarrow & bcZXaa \mid bcX'ZXaa \mid bcZXaaY' \mid bcX'ZXaaY' \\
Z & \rightarrow & c \mid cZ' \\
Z' & \rightarrow & Zb \mid ZbZ'
\end{array}$$

Nyní můžeme přikročit k samotnému převodu na GNF. Uspořádání, požadované v algoritmu, je například toto:

$$Z' \prec Y' \prec X' \prec S \prec X \prec Y \prec Z$$

Dále pokračujeme v substitucích podle algoritmu (následující řádky jsou uspořádány podle toho, v jakém pořadí algoritmus zpracovává jednotlivé neterminály).

$$\begin{array}{rcl}
Z & \rightarrow & c \mid cZ' \\
Y & \rightarrow & aa \mid aaY' \\
X & \rightarrow & aabc \mid aaY'bc \mid aabcX' \mid aaY'bcX' \\
S & \rightarrow & aabcZX \mid aaY'bcZX \mid aabcX'ZX \mid aaY'bcX'ZX \\
X' & \rightarrow & bc \mid bcX' \\
Y' & \rightarrow & bcZXaa \mid bcX'ZXaa \mid bcZXaaY' \mid bcX'ZXaaY' \\
Z' & \rightarrow & cb \mid cZ'b \mid cbZ' \mid cZ'bZ'
\end{array}$$

Nakonec nahradíme terminály na jiných než počátečních pozicích příslušnými neterminály. Výsledná gramatika je tedy $G' = (\{S, X, X', Y, Y', Z, Z', a', b', c'\}, \{a, b, c\}, P', S)$, kde

$$\begin{array}{rrcl}
P = \{ & S & \rightarrow & aa'b'c'ZX \mid aa'Y'b'c'ZX \mid aa'b'c'X'ZX \mid aa'Y'b'c'X'ZX, \\
& X & \rightarrow & aa'b'c' \mid aa'Y'b'c' \mid aa'b'c'X' \mid aa'Y'b'c'X', \\
& X' & \rightarrow & bc' \mid bc'X', \\
& Y & \rightarrow & aa' \mid aa'Y', \\
& Y' & \rightarrow & bc'ZXa'a' \mid bc'X'ZXa'a' \mid bc'ZXa'a'Y' \mid bc'X'ZXa'a'Y', \\
& Z & \rightarrow & c \mid cZ', \\
& Z' & \rightarrow & cb' \mid cZ'b' \mid cb'Z' \mid cZ'b'Z', \\
& a' & \rightarrow & a, \\
& b' & \rightarrow & b, \\
& c' & \rightarrow & c \ \ \}.
\end{array}$$

IB102 – úkol 10 – řešení Odevzdání: 6. 12. 2010

Vypracoval: James Bond UČO: 007
Skupina: MI6

2. [**3 body**] O následujících jazycích rozhodněte, zda jsou bezkontextové, a své rozhodnutí dokažte. (V případě, že jazyk je bezkontextový, nám bude jako důkaz stačit napsání příslušné bezkontextové gramatiky nebo zásobníkového automatu.)

(a) $L_1 = \{w \mid w \in \{a,b\}^+, \#_a(w) = \#_b(w)\}^2$

(b) $L_2 = \{w^2 \mid w \in \{a,b\}^+, \#_a(w) = \#_b(w)\}$

Řešení: Jazyk L_1 je bezkontextový, jazyk L_2 není bezkontexový.
Ad (a): Stačí si uvědomit, že se jazyk L_1 dá ekvivalentně napsat takto:

$$L_1 = \{u \cdot v \mid u, v \in \{a,b\}^+, \#_a(u) = \#_b(u), \#_a(v) = \#_b(v)\}$$

Jedná se tedy o jazyk, jehož slova jsou vždy zřetězením dvou neprázdných slov, z nichž každé má stejný počet a jako b. Tento jazyk generuje např. gramatika $G = (\{S, X\}, \{a, b\}, P, S)$, kde

$$\begin{array}{rlll} P = \{ & S & \rightarrow & XX, \\ & X & \rightarrow & XX \mid aXb \mid bXa \mid ab \mid ba \quad \}. \end{array}$$

Ad (b): Jazyk L_2 je jazyk všech slov, která se dají rozdělit na dvě *stejné* části, které mají stejný počet a jako b. (Jazyk L_2 se tedy od jazyka L_1 značně liší, např. $abba \in L_1$, zatímco $abba \notin L_2$.) Že je jazyk L_2 neregulární, dokážeme pomocí Pumping lemmatu pro bezkontextové jazyky.

- Nechť n je libovolné, nadále pevné.
- Zvolíme slovo $z \in L$, $|z| > n$ takto: $z = a^n b^n a^n b^n$.
- Všechna možná rozdělení $z = uvwxy$ taková, že $vx \neq \varepsilon$, $|vwx| \leq n$ jsou jednoho z těchto druhů:
  - $vwx \in \{a\}^+$. Volbou $i = 0$ pak slovo uv^0wx^0y je buď tvaru $a^m b^n a^n b^n$ nebo tvaru $a^n b^n a^m b^n$, kde $m < n$. Tato slova nepatří do L_2.
  - $vwx \in \{b\}^+$. Podobně jako v předchozím bodě volbou $i = 0$ bude slovo uv^0wx^0y buď tvaru $a^n b^m a^n b^n$ nebo tvaru $a^n b^n a^n b^m$, kde $m < n$. Tato slova opět nepatří do L_2.
  - $vwx \in \{a\}^+ \cdot \{b\}^+$. Pak buď $u \in \{a\}^*$ (vwx je v první polovině slova z) nebo $y \in \{b\}^*$ (vwx je v druhé polovině slova z), obě možnosti zároveň nastat nemohou. Zvolíme opět $i = 0$, v prvním případě pak bude slovo uv^0wx^0y tvaru $a^k b^l a^n b^n$, v druhém případě tvaru $a^n b^n a^k b^l$, kde v obou případech alespoň jedno z l, k je číslo menší než n. Žádné takové slovo nepatří do L_2.

  - $vwx \in \{b\}^+ \cdot \{a\}^+$ (tj. vwx je na rozhraní dvou polovin slova z). Znovu volíme $i = 0$. Slovo uv^0wx^0y pak je určitě tvaru $a^nb^ra^sb^n$, kde alespoň jedno z r, s je menší než n, a tudíž nepatří do L_2.

- Jiná rozdělení nejsou. Ukázali jsme tedy, že pro každé rozdělení je možno najít konstantu i takovou, že $uv^iwx^iy \notin L_2$. Podle PL pro bezkontextové jazyky tedy L_2 není bezkontextový. $\square$