Bezkontextová gramatika - formálně (připomenutí)

Bezkontextová gramatika $G = (\Pi, \Sigma, S, P)$:
Π ... konečná množina *neterminálů*,
Σ ... konečná množina *terminálů* $\quad (\Pi \cap \Sigma = \emptyset)$,
$S \in \Pi$ ... *počáteční neterminál*
P ... konečná množina pravidel typu $A \to \beta$, kde $A \in \Pi$, $\beta \in (\Pi \cup \Sigma)^*$.
Relace $\Rightarrow$ (resp. $\Rightarrow_G$) na $(\Pi \cup \Sigma)^*$:

$$\gamma \Rightarrow \delta$$

jestliže

$$(\exists \mu_1, \mu_2, A, \beta :) \;\; \gamma = \mu_1 A \mu_2,\, \delta = \mu_1 \beta \mu_2,\, (A \to \beta) \in P$$

Relace $\Rightarrow^*$, reflexivní a tranzitivní uzávěr relace $\Rightarrow$:

- jestliže $\alpha \Rightarrow \beta$, pak $\alpha \Rightarrow^* \beta$.
- $\alpha \Rightarrow^* \alpha$;
- jestliže $\alpha \Rightarrow^* \beta$ a $\beta \Rightarrow^* \gamma$, pak $\alpha \Rightarrow^* \gamma$.

Redukce bezkontextových gramatik

$G = (\Pi, \Sigma, S, P)$:

$$
\begin{aligned}
S &\longrightarrow aSb \mid aAbb \mid \varepsilon \\
A &\longrightarrow aAB \mid bB \\
B &\longrightarrow aAb \mid BB \\
C &\longrightarrow CC \mid cS
\end{aligned}
$$

Definujeme induktivně $\mathcal{D}$:

- $S \in \mathcal{D}$;
- $(\, X \in \mathcal{D}, (X \longrightarrow \alpha Y \beta) \in P, Y \in \Pi \,) \Longrightarrow Y \in \mathcal{D}$.

Tedy $\mathcal{D} = \{\, Z \in \Pi \mid \exists \alpha, \beta : S \Rightarrow^* \alpha Z \beta \,\}$.
Definujeme induktivně $\mathcal{T}$:

- $(\, (X \longrightarrow \alpha) \in P, \alpha \in (\Sigma \cup \mathcal{T})^* \,) \Longrightarrow X \in \mathcal{T}$.

Tedy $\mathcal{T} = \{\, Z \in \Pi \mid \exists u \in \Sigma^* : Z \Rightarrow^* u \,\}$.

Věta. Ke každé BG G, pro niž $L(G) \neq \emptyset$, lze sestrojit redukovanou gramatiku G' tž. $L(G') = L(G)$. (Jak?)

Chomského normální forma bezkontextových gramatik

Věta. Ke každé BG G lze sestrojit nevypouštějící gramatiku G' (tj. bez pravidel typu $X \longrightarrow \varepsilon$) tž. $L(G') = L(G) - \{\varepsilon\}$.

BG je v *Chomského normální formě*, jestliže její pravidla jsou výhradně ve tvarech $X \longrightarrow YZ$ a $X \to a$.
Věta. Ke každé BG G lze sestrojit BG G' v ChNF tž. $L(G') = L(G) - \{\varepsilon\}$.
Převod gramatiky G do ChNF se dá rozdělit do čtyř kroků:

- převod na nevypouštějící gramatiku,
- odstranění pravidel typu $X \to Y$,
- pro každý terminál a přidáme nový neterminál A_a a pravidlo $A_a \to a$; na každé pravé straně delší než 1 nahradíme a neterminálem A_a,
- každé pravidlo typu $X \to Y_1 Y_2 \ldots Y_n$, kde $n \geq 3$, nahradíme soustavou pravidel využívající nově přidané neterminály.

Příklad.

$$
\begin{aligned}
S &\longrightarrow (E) \mid E \\
E &\longrightarrow F + F \mid F \times F \\
F &\longrightarrow a \mid S
\end{aligned}
$$

Zásobníkové automaty

(Nedeterministický) zásobníkový automat ... $M = (Q, \Sigma, \Gamma, \delta, q_0, Z_0)$, kde Q je konečná množina *stavů*, Σ je konečná *vstupní abeceda*, Γ je konečná *zásobníková abeceda*, $q_0 \in Q$ je *počáteční stav*, $Z_0 \in \Gamma$ je *počáteční zásobníkový symbol* a

$$\delta : Q \times (\Sigma \cup \{\varepsilon\}) \times \Gamma \to \mathcal{P}_{fin}(Q \times \Gamma^*)$$

je přechodová funkce (neboli konečná množina instrukcí).
Konfigurace ZA M: (q, w, α), kde $q \in Q$, $w \in \Sigma^*$, $\alpha \in \Gamma^*$.
Relace $\vdash_M$ (odvození v jednom kroku) na množině konfigurací:

$$(q, aw, X\beta) \vdash_M (q', w, \alpha\beta) \Leftrightarrow \delta(q, a, X) \ni (q', \alpha)$$

(kde $a \in (\Sigma \cup \{\varepsilon\})$, $w \in \Sigma^*$, $\beta \in \Gamma^*$).
Relace $\vdash_M^*$ (odvození v konečně mnoha krocích) je reflexivním a tranzitivním uzávěrem relace $\vdash_M$.
Jazyk rozpoznávaný (nebo též *přijímaný*) ZA M:

$$L(M) = \{\, w \in \Sigma^* \mid (q_0, w, Z_0) \vdash_M^* (q, \varepsilon, \varepsilon) \text{ pro nějaký } q \in Q \,\}.$$

Deterministické zásobníkové automaty

ZA $M = (Q, \Sigma, \Gamma, \delta, q_0, Z_0)$ je *deterministický*, jestliže

- $\delta(q, a, X)$ je vždy nejvýše jednoprvková množina (pro $a \in \Sigma \cup \{\varepsilon\}$) (tedy neexistují dvě různé instrukce se stejnou levou stranou),
- je-li $\delta(q, \varepsilon, X) \neq \emptyset$, pak $\delta(q, a, X) = \emptyset$ pro vš. $a \in \Sigma$ (může-li automat udělat ε-krok, nemůže mu jiná instrukce zároveň umožňovat přečtení vstupního symbolu).

Pro jazyk

$$L = \{wcw^R \mid w \in \{a, b\}^*\}$$

snadno sestrojíme deterministický zásobníkový automat (přijímající L prázdným zásobníkem).
Snadnou úpravou dostaneme nedeterministický ZA pro jazyk

$$L' = \{ww^R \mid w \in \{a, b\}^*\}$$

(Pro L' neexistuje deterministický přijímající ZA).

Přijímání jazyka prázdným zásobníkem a přijímajícími stavy

$M = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$, kde $F \subseteq Q$

$L_{PZ}(M) = \{\, w \in \Sigma^* \mid (q_0, w, Z_0) \vdash_M^* (q, \varepsilon, \varepsilon)$ pro nějaký $q \in Q \,\}$

$L_{KS}(M) = \{\, w \in \Sigma^* \mid (q_0, w, Z_0) \vdash_M^* (q, \varepsilon, \alpha)$ pro nějaké $q \in F, \alpha \in \Gamma^* \,\}$

Tvrzení:
Ke každému (nedet.) ZA M lze (snadno) sestrojit ZA M' tž.
$L_{KS}(M') = L_{PZ}(M)$.
Naopak také: Ke každému (nedet.) ZA M lze (snadno) sestrojit ZA M' tž.
$L_{PZ}(M') = L_{KS}(M)$.

Tvrzení:
Ke každému DZA M lze (snadno) sestrojit DZA M' tž.
$L_{KS}(M') = L_{PZ}(M)$.
Ke každému (nedet.) ZA M lze (snadno) sestrojit ZA M' tž.
$L_{PZ}(M') = L_{KS}(M) \cdot \{\$\}$. (Je nutné přidat speciální koncový symbol, např. \$, který není v původní abecedě. Prázdným zásobníkem totiž může DZA přijímat jen bezprefixové jazyky.)

Zásobníkové automaty a bezkontextové jazyky

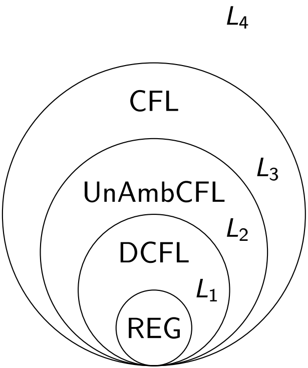
Věta. (Nedeterministické) zásobníkové automaty rozpoznávají právě bezkontextové jazyky, tedy jazyky z třídy CFL.

BG $\longrightarrow$ ZA

Pro BG $G = (\Pi, \Sigma, S, P)$ *sestrojme*
$M = (\{q_0\}, \Sigma, \Pi \cup \Sigma, \delta, q_0, S)$, *kde*
pro $X \in \Pi$: $\delta(q_0, \varepsilon, X) = \{\, (q_0, \alpha) \mid (X \to \alpha) \in P \,\}$
a pro $a \in \Sigma$: $\delta(q_0, a, a) = \{\, (q_0, \varepsilon) \,\}$.
(Všude jinde δ *přiřazuje* $\emptyset$*.)*

1/ $A \longrightarrow A + B$
2/ $A \longrightarrow B$
3/ $B \longrightarrow B * C$
4/ $B \longrightarrow C$
5/ $C \longrightarrow (A)$
6/ $C \longrightarrow a$

Hierarchie jazyků



$$L_4 = \{a^n b^n c^n \mid n \geq 0\}$$

$$L_3 = \{a^i b^j c^k \mid i = j \vee j = k\}$$

$$L_2 = \{ww^R \mid w \in \{a, b\}^*\}$$

$$L_1 = \{a^n b^n \mid n \geq 1\}$$

Uzávěrové vlastnosti třídy CFL

Snadnými konstrukcemi se prokáže, že CFL je uzavřena např. vůči těmto operacím:

- sjednocení ($S \longrightarrow S_1 \mid S_2$)
- zřetězení ($S \longrightarrow S_1 S_2$)
- iterace ($S \longrightarrow \varepsilon \mid S_1 S$)
- zrcadlový obraz ($A \longrightarrow \alpha \ \ldots \ A \longrightarrow (\alpha)^R$)

CFL ale není uzavřena např. vůči

- průniku ($\{a^i b^j c^k \mid i = j\} \cap \{a^i b^j c^k \mid j = k\} = \{a^n b^n c^n \mid n \geq 0\}$)
- doplňku (z předchozího díky de Morganovým pravidlům; konkrétně např. doplňky jazyků $\{a^n b^n c^n \mid n \geq 0\}$ či $\{ww \mid w \in \{0,1\}^*\}$ jsou bezkontextové, ale doplňky těchto doplňků [tedy tyto jazyky] bezkontextové nejsou)