Teorie jazyků a automatů— 8. týden

Marie Demlová
http://math.feld.cvut.cz/en/people/demlova

11. 11. 2024

Gramatiky

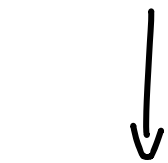
Nevypouštěcí gramatiky.

Gramatiku $\mathcal{G} = (N, \Sigma, S, P)$ nazveme nevypouštěcí, jestliže neobsahuje žádné pravidlo typu $A \to \varepsilon$.

Tvrzení.

Ke každé CF $\mathcal{G}$ existuje nevypouštěcí gramatika $\mathcal{G}_1$ taková, že

$$L(\mathcal{G}_1) = L(\mathcal{G}) - \{\varepsilon\}.$$



bezkontextová

2d. $V = \{A \mid A \Rightarrow^* \varepsilon\}$

$A \Rightarrow \varepsilon \quad B \to AA$

$A \to \varepsilon \quad A \to \varepsilon$

$B \Rightarrow A \quad A \Rightarrow A \Rightarrow \varepsilon$

$V_1 = \{A \mid A \to \varepsilon \in P\}$

$V_2 = V_1 \cup \{A \mid A \to \alpha \in P, \alpha \in V_1^*\}$

$V_{i+1} = V_i \cup \{A \mid A \to \alpha \in P, \alpha \in V_i^*\}$

$V_1 \subseteq V_2 \subseteq V_3 \subseteq \ldots \subseteq V_i \ldots \subseteq N$

tj. ex. i $V_i = V_{i+1} = \underline{V}$

Gramatiky

Příklad.

Dána gramatika $\mathcal{G} = (\{A, S\}, \{a, b, c\}, S, P)$ pravidly

$$S \to aSc \,|\, A \quad A \to bAc \,|\, \varepsilon.$$

Ke gramatice $\mathcal{G}$ najdeme nevypouštěcí gramatiku $\mathcal{G}_1$, která generuje jazyk $L(\mathcal{G}) - \{\varepsilon\}$.

G_1 : z P odstraníme pravidla pro $A \to \varepsilon$

$V_0 = \{A\}$

$V_1 = \{A\} \cup \{S\} = \{A, S\} = N$

pro každé pravidlo γ

$A \to \alpha$

$A \to \beta_1 X_1 \beta_2 X_2 \dots \beta_k X_k \beta_{k+1}, \quad X_i \in V$

β_i nemají neterminál z V

pro každou $M \subseteq \{X_1, \dots, X_k\}$

$A \to \alpha'$ kde α' jsme vynechali $X_i \in M$

G_1: $S \to aSC \mid acdA$

$A \to bAc \mid bc$

Gramatiky

Důsledek.

Označme $\mathcal{L}_i$ třídu jazyků typu i. Pak platí:

$$\mathcal{L}_3 \subseteq \mathcal{L}_2 \subseteq \mathcal{L}_1 \subseteq \mathcal{L}_0.$$

Tvrzení 1. Je-li L regulární jazyk, pak existuje regulární gramatika $\mathcal{G}$ taková, že $L = L(\mathcal{G})$.

Tvrzení 2. Je-li $\mathcal{G}$ gramatika typu 3, pak jazyk $L(\mathcal{G})$ je regulární.

$L \in Reg$ tj. ex. DFA $M = (Q, \Sigma, \delta, q_0, F)$ $\overset{Q \cap E = \emptyset}{\rightsquigarrow}$ $G = (N, \Sigma, S, P)$

$N = Q \quad S = q_0 \quad \delta(q, a) = p \quad F \subseteq Q$

$q \to ap \quad q_f \to \varepsilon \ \forall q_A \in F$

$w \in L(M)$ iff $\delta^*(q_0, w) = q_A \in F$

$w = a_1 \dots a_k \quad q_0, \ \delta(q_0, a_1) = q_1, \ \delta^*(q_0, a_1 a_2) = \delta(q_1, a_2) = q_2 \quad q_A$

$q_0 \to a_1 q_1 \quad q_1 \to a_2 q_2 \quad \dots \quad q_{k-1} \to a_k q_k$

$q_0 \Rightarrow a_1 q_1 \Rightarrow a_1 a_2 q_2 \Rightarrow \dots \Rightarrow a_1 \dots a_k q_f \overset{q_f \to \varepsilon}{\Rightarrow} a_1 \dots a_k$

$a_1 \dots a_k = w \in L(G)$

$q_0 \overset{*}{\Rightarrow} n q_f \Rightarrow n$ pro něj $q_f \in F$

G typu 3 P $A \to wB$ nebo $A \to w$

$w \in \Sigma^*$

$G = (N, \Sigma, S, P)$ $A, B \in N$ $|w| = 0$ nebo $|w| = 1$

$|w| > 1$ $A \to a_1 a_2 \dots a_k B$ přidáme nové neterminály $B_1, \dots, B_{k-1}$

$A \to a_1 B_1$ $B_1 \to a_2 B_2$ $B_{k-2} \to a_{k-1} B_{k-1}$ $B_{k-1} \to a_k B$

$A \to b_1 C_1$ $C_1 \to b_2 C_2$ $\dots$ $C_{n-1} \to b_n C_n$ $C_n \to \varepsilon$

$M = (Q, \Sigma, \delta, q_0, F)$ $Q = N$ $q_0 = S$

$A = bB$ iff $B \in \delta(A, b)$

$F = \{A \mid A \to \varepsilon \in P\}$

Bezkontextové gramatiky

Tvrzení.

Dána CF gramatika $\mathcal{G} = (N, \Sigma, S, P)$ a v ní derivace

$$S \Rightarrow_{\mathcal{G}}^{\star} \alpha A \beta \Rightarrow_{\mathcal{G}}^{\star} \gamma,$$

pro $\alpha, \beta, \gamma \in (\Sigma \cup N)^{\star}$ a $A \in N$.
Pak existují slova $\varphi, \psi, \mu \in (\Sigma \cup N)^{\star}$ taková, že

$$\gamma = \varphi\psi\mu \quad \text{a} \quad \alpha \Rightarrow_{\mathcal{G}}^{\star} \varphi, \quad A \Rightarrow_{\mathcal{G}}^{\star} \psi, \quad \beta \Rightarrow_{\mathcal{G}}^{\star} \mu.$$

$S \Rightarrow^* \alpha A \beta \Rightarrow^* \gamma$ $\quad \gamma = \varphi\psi\mu$ tak, že $\alpha \Rightarrow^* \varphi$, $A \Rightarrow^* \psi$, $\beta \Rightarrow^* \mu$

Př. $S \to aAAb$

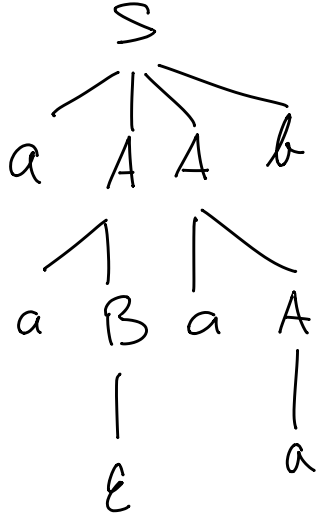
$A \to aB \mid aA \mid a$

$B \to bB \mid \varepsilon$

$$S \Rightarrow aAAb \overset{A\to aA}{\Longrightarrow} aAaAb \overset{A\to aB}{\Longrightarrow} aaBaAb \overset{A\to a}{\Longrightarrow} aaBaab \overset{B\to\varepsilon}{\Longrightarrow} aaaab$$

$$S \Rightarrow aAAb \overset{A\to aB}{\Longrightarrow} aaBAb \overset{B\to\varepsilon}{\Longrightarrow} aaAb \overset{A\to aA}{\Longrightarrow} aaaAb \overset{A\to a}{\Longrightarrow} aaaab$$

$\longrightarrow$ stejná derivace s jiným pravidlem (levá)



Bezkontextové gramatiky

Levá derivace, levé odvození

Přímé odvození je levé, jestliže se přepisuje ten neterminál, který je nejvíc „vlevo“.

Derivace (odvození) je levá derivace (levé odvození), jestliže se skládá pouze z levých přímých odvození.

Tvrzení. Je dána bezkontextová gramatika $\mathcal{G} = (N, \Sigma, S, P)$. Pak pro každou derivaci $S \Rightarrow_{\mathcal{G}}^{\star} w$ existuje levá derivace terminálního w z S taková, že používá stejná pravidla jako původní derivace (pouze možná v jiném pořadí).

Bezkontextové gramatiky

Derivační strom (parse tree)

Dána CF gramatika $\mathcal{G} = (N, \Sigma, S, P)$. Derivační strom (anglicky parse tree) je kořenový strom, takový, že:

1. Každý vrchol, který není list, je ohodnocen neterminálem.
2. Každý list je ohodnocen neterminálem, terminálem nebo prázdným slovem ε. V případě, že je list ohodnocen prázdným slovem ε, je to jediný následník (svého předchůdce).
3. Jestliže některý vrchol, který není list, je ohodnocen neterminálem A a má následníky (v pořadí zdleva doprava) $X_1, X_2, \ldots, X_k$, $X_i \in N \cup \Sigma$, pak $A \to X_1 X_2 \ldots X_k$ je pravidlo gramatiky $\mathcal{G}$.

Řekneme, že derivační strom dává, nebo má za výsledek slovo α, jestliže α je ohodnocení listů derivačního stromu (čteno zleva doprava).

Bezkontextové gramatiky

Tvrzení.

1. Pro každou derivaci $S \Rightarrow_{\mathcal{G}}^{\star} \alpha$ existuje derivační strom s výsledkem α.
2. Ke každému derivačnímu stromu s výsledkem α existuje aspoň jedna derivace $S \Rightarrow_{\mathcal{G}}^{\star} \alpha$ (takových derivací může být více).
3. Ke každému derivačnímu stromu s výsledkem α existuje právě jedna levá (právě jedna pravá) derivace α z S.

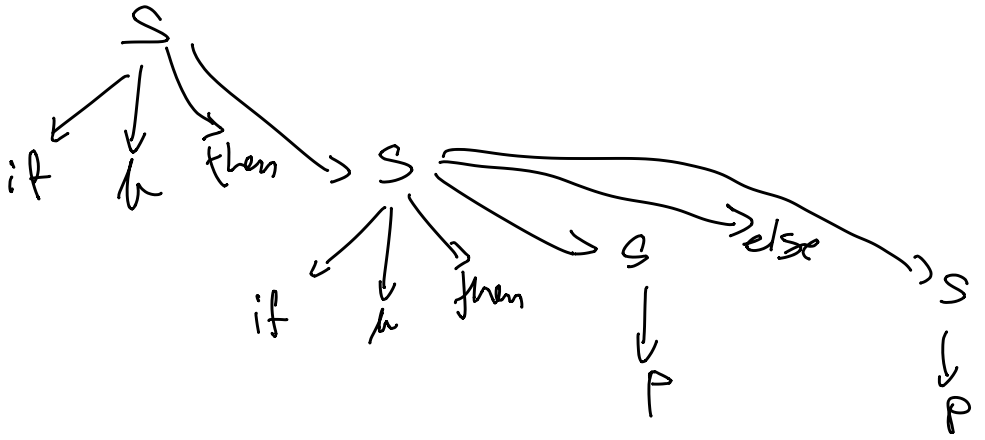
Bezkontextové gramatiky

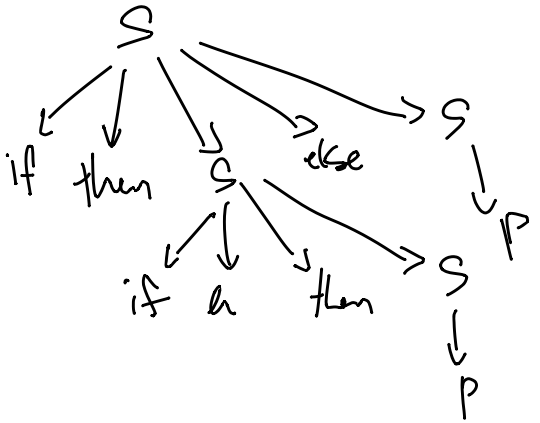
Jednoznačné a víceznačné bezkontextové gramatiky.

CF gramatika $\mathcal{G} = (N, \Sigma, S, P)$ je jednoznačná, jestliže pro každé slovo w generované $\mathcal{G}$ existuje jediný derivační strom s výsledkem w (tj. existuje jediná levá derivace w z S). V opačném případě mluvíme o víceznačné gramatice.

$S \to$ if b then S else S | if b then S | p ambiguous grammar

if b then if b then p else p





fix:

$S \to$ if b then S | if b then S_1 else S | p

$S_1 \to$ if b then S_1 else S_1 | p

Bezkontextové gramatiky

Víceznačný jazyk.

Jazyk L je víceznačný, jestliže **každá** bezkontextová gramatika, která ho generuje, je víceznačná.

Například jazyk $L = \{a^i b^j c^k d^l \mid i = j \text{ nebo } k = l\}$ je víceznačný.

Redukovaná bezkontextová gramatika

CF gramatika $\mathcal{G} = (N, \Sigma, S, P)$, která generuje aspoň jedno slovo, je redukovaná, jestliže splňuje tyto dvě podmínky:

1. Ke každému neterminálu A existuje aspoň jedno terminální slovo w takové, že $A \Rightarrow_{\mathcal{G}}^{\star} w$.
2. Ke každému neterminálu A existují slova $\alpha, \beta \in (N \cup \Sigma)^{\star}$ tak, že
$$S \Rightarrow_{\mathcal{G}}^{\star} \alpha A \beta.$$

Tvrzení. Ke každé CF gramatice $\mathcal{G} = (N, \Sigma, S, P)$, pro kterou $L(\mathcal{G}) \neq \emptyset$, existuje redukovaná gramatika $\mathcal{G}_1$ taková, že $L(\mathcal{G}_1) = L(\mathcal{G})$.

1) $\forall A \in N$ existuje $w \in \Sigma^*$, že $A \Rightarrow^* w$

2) $\forall A \in N$ existují $\alpha, \beta \in (N \cup \Sigma)^*$, že $S \Rightarrow^* \alpha A \beta$

$S \Rightarrow \alpha A \beta \Rightarrow^* w \in \Sigma^*$

$S \to SCA \mid a$

$A \to aCb \mid \varepsilon$

$B \to bC \mid BA \mid BD$

$C \to AA \mid b$

$D \to ABC$

$V_1 = \{S, A, C\}$ ($a \to$, $A \to \varepsilon$, $C \to b$)

$V_2 = \{S, A, C\} \cup \{B\}$ $B \to bC$

$V_3 = \{S, A, C, B\} \cup \{D\}$ $V_3 = N$

1) $V = \{A \mid A \Rightarrow^* w, w \in \Sigma^*\}$

$V_1 = \{A \mid A \Rightarrow w \in P, w \in \Sigma^*\}$

$V_2 = V_1 \cup \{A \mid A \to \alpha \in P, \alpha \in (\Sigma \cup V_1)^*\}$

$V_{i+1} = V_i \cup \{A \mid A \to \alpha \in P, \alpha \in (\Sigma \cup V_i)^*\}$

$V_1 \subseteq V_2 \subseteq \ldots \subseteq V_i \subseteq N \quad V_i = V_{i+1} = V$

$L(G) \neq \emptyset$ iff $S \in V$

2) $V \quad U = \{A \mid A \in V, S \Rightarrow^* \alpha A \beta\}$

$U_0 = \{S\} \quad U_{i+1} = U_i \cup \{A \mid B \to \alpha A \beta \quad B \in U_i\}$

$U_0 = \{S\}$

$U_1 = \{S\} \cup \{C, A\}$

$U_2 = \{S, S, A\} \cup \emptyset = U$

$S \to SCA \mid a$

$A \to aCb \mid \varepsilon$

$C \to AA \mid b$

Redukovaná bezkontextová gramatika

Algoritmus redukce.

Dána CF gramatika $\mathcal{G} = (N, \Sigma, S, P)$.

1. Sestrojíme množinu $V = \{A \mid A \in N, A \Rightarrow_{\mathcal{G}}^{\star} w, w \in \Sigma^{\star}\}$

   Jestliže $S \notin V$, pak $L(\mathcal{G}) = \emptyset$ a redukovaná gramatika ke gramatice $\mathcal{G}$ neexistuje.

   Definujeme $\mathcal{G}' = (V, \Sigma, S, P')$: do P' dáme ta pravidla z P, která obsahují pouze neterminály z množiny V.

Redukovaná bezkontextová gramatika

2 Pro gramatiku $\mathcal{G}' = (V, \Sigma, S, P')$ sestrojíme (opět indukcí) množinu

$$U = \{A \mid A \in V, \text{ existují } \alpha, \beta \in (V \cup \Sigma)^\star \text{ tak, že } S \Rightarrow_{\mathcal{G}}^\star \alpha A \beta\}.$$

Hledaná gramatika je $\mathcal{G}_1 = (U, \Sigma, S, P_1)$, kde P_1 je množina všech pravidel z P' (a tedy i z P), které obsahují neterminály pouze z množiny U.

Redukovaná bezkontextová gramatika

Příklad

Je dána bezkontextová gramatika $\mathcal{G} = (N, \Sigma, S, P)$, kde $N = \{S, A, B, C, D\}$, $\Sigma = \{a, b\}$ a P je dáno

$$
\begin{aligned}
S &\to SCA \mid a \\
A &\to aCb \mid \varepsilon \\
B &\to bC \mid BA \mid BD \mid \varepsilon \\
C &\to AA \mid b \\
D &\to ABC
\end{aligned}
$$

Danou gramatiku redukujte.