Teorie jazyků a automatů— 9. týden

Marie Demlová

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

18. 11. 2024

Bezkontextové gramatiky

kontextová gramatika $\alpha X \beta \to \alpha \gamma \beta$ $\alpha, \beta, \gamma \in (\Sigma \cup N)^*$
$X \in N$

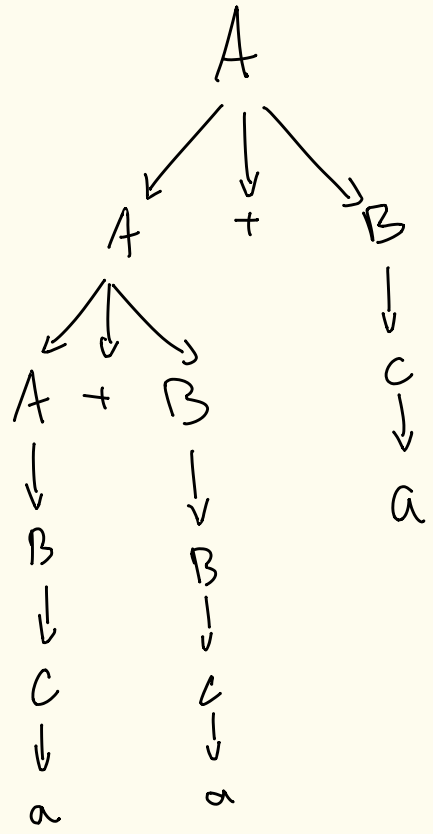
Příklady

1. Gramatika pro aritmetické výrazy. $G = (N, \Sigma, S, P)$ $N = \{A, B, C\}$ $S = A$ $\Sigma = \{+, *, (,), a\}$

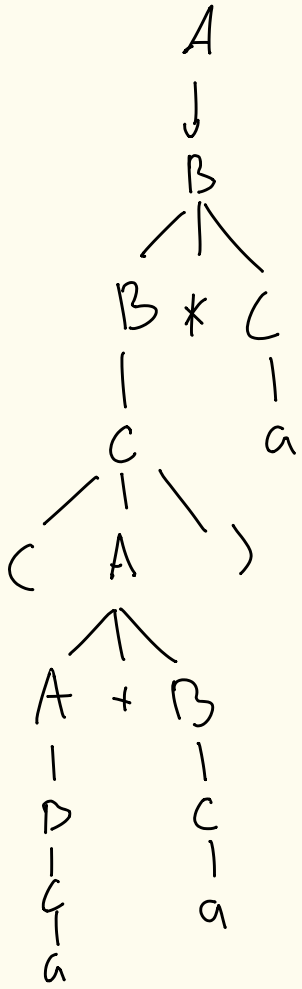
   P. $A \to A+B \mid B$
   $B \to B*C \mid C$
   $C \to (A) \mid a$

2. Gramatika pro regulární výrazy.

$a+a+a$



$(a+a)*a$



(2) Gramatika reg výrazů $\{a,b,c\}$

$N=\{A,B,C,D\}$ $\Sigma=\{a,b,c,+,*,(,)\}$

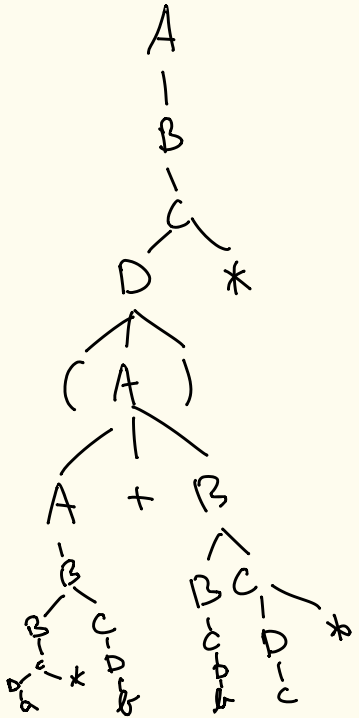
$A \to A+B \mid B$

$B \to BC \mid C$

$C \to D* \mid D$

$D \to (A) \mid a \mid b \mid c$

$(a^*b+bc^*)^*$



Chomského normální tvar

Gramatika $\mathcal{G}$ je v Chomského normálním tvaru, jestliže má pouze pravidla tvaru

$$A \to BC,\ A \to a \quad \text{pro} \quad A, B, C \in N,\ a \in \Sigma.$$

Věta. Pro každou CF gramatiku $\mathcal{G}$ existuje CF gramatika $\mathcal{G}'$ v Chomského normálním tvaru taková, že

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

Chomského normální tvar

Postup nalezení ekvivalentní gramatiky v Chomského normálním tvaru.

1. Sestrojíme nevypouštěcí gramatiku $\mathcal{G}_1$, která generuje všechna neprázdná slova z jazyka $L(\mathcal{G})$.
2. V $\mathcal{G}_1$ odstraníme pravidla typu $A \to B$, kde $A, B \in N$.
3. Odstraníme terminály $a \in \Sigma$, které se vyskytují na pravé straně pravidla, které má délky aspoň 2.
4. Zkrátíme pravé strany pravidel, které jsou delší než 2.

Chomského normální tvar

Příklad

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

$$S \to SCA \mid a$$
$$A \to aCb \mid \varepsilon$$
$$C \to AA \mid b$$

Danou gramatiku převeďte do Chomského normálního tvaru.

(1) Nevypouštěcí: $V_1 = \{A\}$ $V_2 = \{A, C\} = V_3 = V$

$C \to AA$

$S \to SCA \mid SA \mid SC \mid a$

$A \to aCb \mid ab$

$C \to AA \mid A \mid b$

(2) $S \to SCA \mid SA \mid SC \mid a$

$A \to aCb \mid ab$

$C \to AA \mid aCb \mid ab \mid b$

(3) $S \to SCA \mid SA \mid SC \mid a$

$A \to X_a C X_b \mid X_a X_b$

$C \to AA \mid X_a C X_b \mid X_a X_b \mid b$

$X_a \to a$

$X_b \to b$

(4) $S \to SB \mid SA \mid SC \mid a$

$B \to CA$

$A \to X_a D \mid X_a X_b$

$C \to AA \mid X_a D \mid X_a X_b \mid b$

$D \to C X_b$

$X_a \to a$

$X_b \to b$

Pumping lemma pro bezkontextové gramatiky

Tvrzení.

Dána CF gramatika $\mathcal{G} = (N, \Sigma, S, P)$ v Chomského normálním tvaru a derivační strom s výsledkem $w \in \Sigma^\star$.

Jestliže nejdelší orientovaná cesta v tomto derivačním stromu má délku n, pak pro délku slova w platí

$$|w| \leq 2^{n-1}.$$

P.L. pro CF jazyky

Když L je CF, tak existuje $m \geq 1$ s vlastností každé $z \in L$, $|z| \geq m$ lze rozložit $z = uvwxy$

(1) $|vwx| \leq m$ (2) $vx \neq \varepsilon$ (3) $uv^i wx^i y \in L$ $\forall i \geq 0$

Př. $L = \{0^n 1^n 2^n \mid n \geq 0\}$ není CF

Kdyby L byl CF, tak existuje $m \geq 1$ s vlastnostmi 1,2,3.

zvolíme $z = 0^m 1^m 2^m \in L$ $\quad |z| = 3m \geq m$

Kdyby $z = uvwxy$, že $|vwx| \leq m$.

Tedy vwx je podslovo $0^m 1^m$ nebo vwx — " — $1^m 2^m$

- $v = 0^i$, $x = 0^j$, $i+j \geq 1$
- $v = 0^i$ $x = 1^j$, $i+j \geq 1$
- $v = 1^i$ $x = 1^j$, $i+j \geq 1$
- $v = 1^i$ $x = 2^j$, $i+j \geq 1$
- $v = 2^i$ $x = 2^j$, $i+j \geq 1$

$0^{m+i+j} 1^m 2^m \notin L$

⋮ ostatní obdobně

D) $m = 2^{|N|}$

$z \in L$

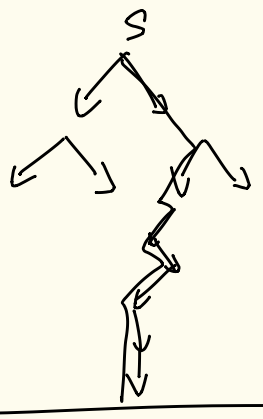
$|z| \geq 2^{|N|} = n$

$|z| \geq 2^n$

$\underbrace{|N|}_{n}+1$

... atd

G v Chomského normálním tvaru



---

délky $n+1$

Pumping lemma pro bezkontextové gramatiky

Věta. Pro každý CF jazyk L existuje přirozené číslo $m \geq 1$ takové, že každé slovo $z \in L$ délky aspoň m lze rozdělit na pět částí $z = uvwxy$ tak, že

- $|vwx| \leq m$, (tj. prostřední část není příliš dlouhá),
- $vx \neq \varepsilon$ (tj. aspoň jedno ze slov v, x není prázdné),
- pro všechna $i \geq 0$ platí $uv^i wx^i y \in L$, (tj. v a x se dají do slova z „napumpovat“ a stále dostaneme slovo z jazyka L).

Pumping lemma pro bezkontextové gramatiky

Využití Pumping lemmatu pro bezkontextové gramatiky.

Příklad. Ukážeme, že jazyk $L = \{0^n 1^n 2^n \mid n \geq 0\}$ není bezkontextový.

Algoritmus CYK

Označme $\mathcal{G} = (N, \Sigma, S, P)$ v Chomského normálním tvaru a $w = a_1 a_2 \dots a_k$.

Vytváříme množiny $X_{i,j}$ pro $1 \leq i \leq j \leq k$, kde

$$X_{i,j} = \{A \in N \mid A \Rightarrow_{\mathcal{G}}^{\star} a_i a_{i+1} \dots a_j\}.$$

Tvrzení. Platí $w \in L(\mathcal{G})$ právě tehdy, když $S \in X_{1,k}$.

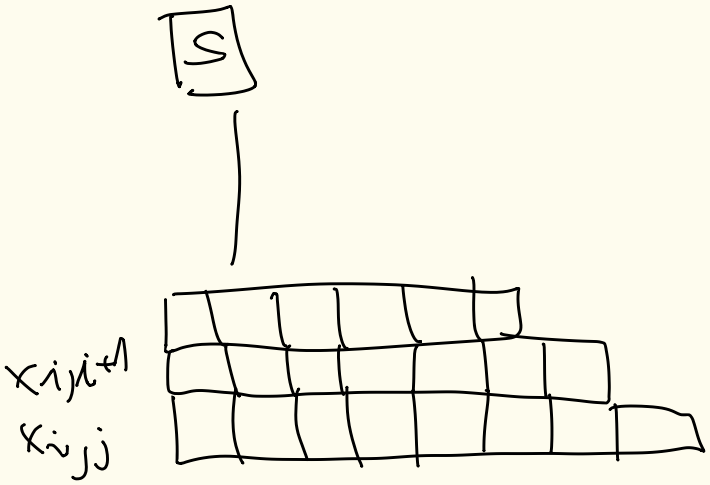
Algoritmus CYK

Příklad. Je dána gramatika $\mathcal{G}$ pravidly

$$\begin{aligned} S &\to AB \mid BC \\ A &\to BA \mid a \\ B &\to CC \mid b \\ C &\to AB \mid a \end{aligned}$$

Pomocí algoritmu CYK rozhodněte, zda slovo *aabab* je generováno bezkontextovou gramatikou $\mathcal{G}$.

$X_{i,i+3}$ $A \to BC \in P$ $\begin{cases} B \in X_{i,i} \; C \in X_{i+1,i+3} \\ B \in X_{i,i+1} \; C \in X_{i+2,i+3} \\ B \in X_{i,i+2} \; C \in X_{i+3,i+3} \end{cases}$

$X_{1,k} = \{A \mid A \Rightarrow^* a_1 \dots a_k\}$

$w \in L(G)$ iff $S \in X_{1,n}$