Teorie jazyků a automatů— 13. týden

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

16. 12. 2024

spojování pravidel $A \to \alpha B \beta \quad B \to \beta_1 | \beta_2 | \dots \beta_k$

$$A \to \alpha \beta_1 \beta | \alpha \beta_2 \beta | \dots | \alpha \beta_k \beta$$

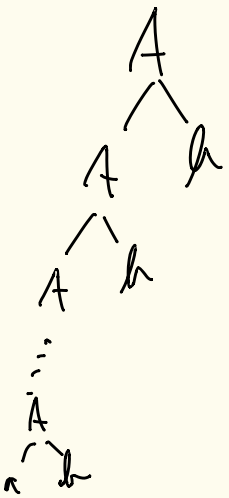
odstranění levé rekurze $A \to A\alpha_1 | A\alpha_2 \dots | A\alpha_k$

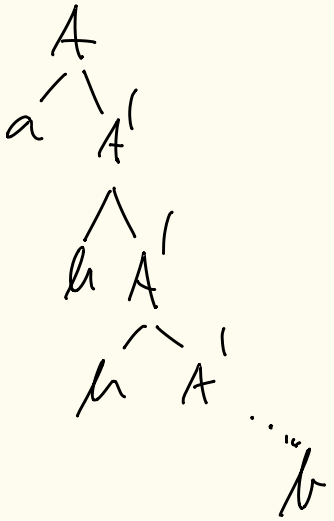
$A \to \gamma_1 | \gamma_2 | \dots | \gamma_l$

$$A \to \gamma_1 | \dots | \gamma_k | \gamma_1 A' | \dots | \gamma_l A'$$

$$A' \to \alpha_1 | \alpha_2 \dots | \alpha_1 A' | \dots | \alpha_k A'$$

$A \to Ab \quad A \to a$





Greibachové normální forma (tvar)

CF gramatika $\mathcal{G}$ je v Greibachové normální formě, jestliže všechna pravidla mají tvar:

$$A \to a\alpha, \quad \text{kde } a \in \Sigma \text{ a } \alpha \in N^{\star}.$$

$A \to aXYZ \quad XYZ \in N^*$

$A \to a \quad a \in \Sigma$

Greibachové normální forma (tvar)

CF gramatika $\mathcal{G}$ je v Greibachové normální formě, jestliže všechna pravidla mají tvar:

$$A \to a\alpha, \quad \text{kde } a \in \Sigma \text{ a } \alpha \in N^\star.$$

Věta. Ke každému CF jazyku L existuje CF gramatika v Greibachové normálním tvaru $\mathcal{G}$ taková, že

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

Greibachové normální forma (tvar)

Postup nalezení gramatiky v Greibachové normálním tvaru.

Máme nevypouštěcí gramatiku $\mathcal{G}$.

(1) Očíslujeme její neterminály, tj. $N = \{A_1, A_2, \ldots, A_n\}$.

(2) Postupným spojováním pravidel a odstraňováním levé rekurze získáme pouze pravidla, kde pravá strana začíná buď terminálem nebo jsou tvaru $A_i \to A_j\alpha$ pro $i < j$.

(3) Zajistíme, aby pravidla obsahující nové neterminály (vzniklé při odstraňování levé rekurze) také měla požadovaný tvar.

Greibachové normální forma (tvar)

Postup nalezení gramatiky v Greibachové normálním tvaru – 2.

(4) Postupným spojováním pravidel (od n do 1) dosáhneme toho, aby pravidla s levou stranou A_i měla tvar $A_i \to a\alpha$, kde $a \in \Sigma$.

(5) Jestliže nyní máme terminální symbol uvnitř pravé strany některého pravidla, zavedeme za něj nový neterminál.

Greibachové normální forma (tvar)

Příklad. Je dána bezkontextová gramatika $\mathcal{G} = (N, \Sigma, S, P)$, kde $N = \{E, T, F\}$, $\Sigma = \{+, \star, (,), a\}$, $S = E$ a

$A_1 = S \; A_2 = T \; A_3 = F$

$$P: \quad \begin{aligned} E &\to E + T \mid T \\ T &\to T \star F \mid F \\ F &\to (E) \mid a \end{aligned}$$

Gramatiku převeďte do Greibachové normálního tvaru.

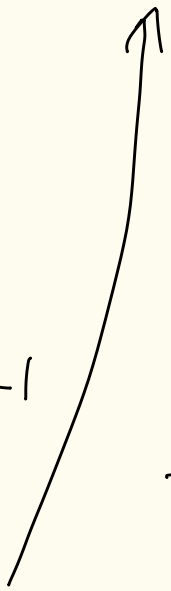
$E \to T \mid TE'$

$E' \to +T \mid +TE'$

$T \to F \mid FT'$

$T' \to *F \mid *FT'$

$F \to (E) \mid a$



$E \to (EX \mid a \mid (EXT' \mid aT' \mid (EXE' \mid aE' \mid (EXT'E' \mid aT'E'$

$E' \to +T \mid +TE'$

$T \to (EX \mid a \mid (EXT' \mid aT'$

$T' \to *F \mid *FT'$

$F = (EX \mid a$

$X \to)$

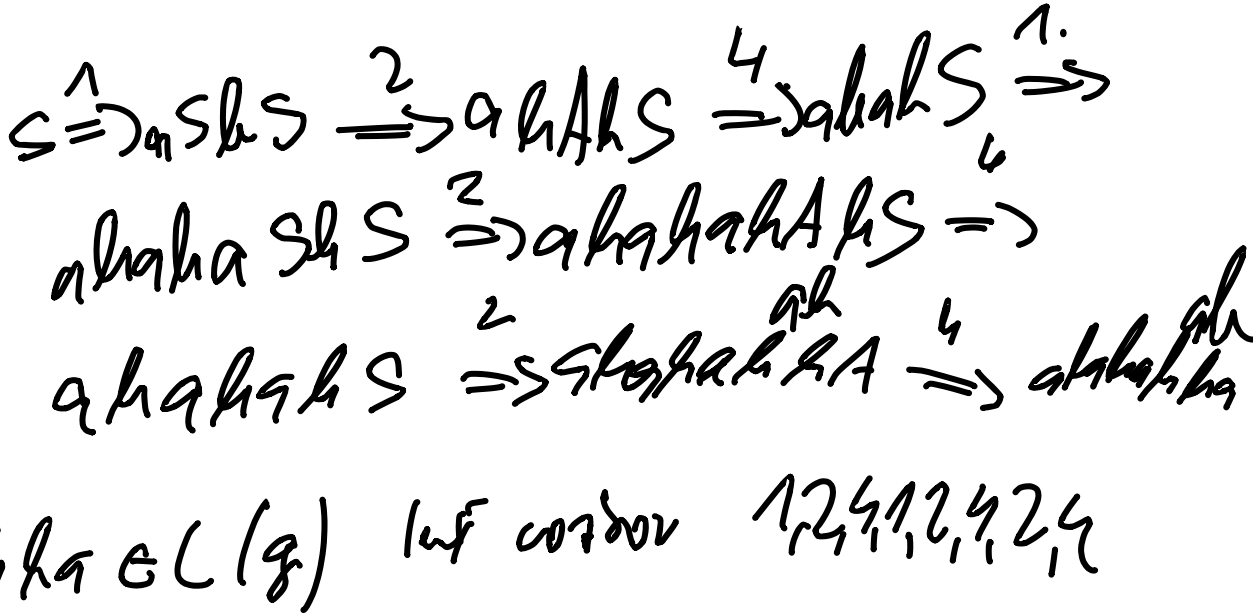
Analýza shora

Příklad. Je dána CF gramatika $\mathcal{G}$ s jedním neterminálem S, $\Sigma = \{a, b\}$ a pravidly:

$$S \to aSbS, S \to bA, A \to bA, A \to a.$$

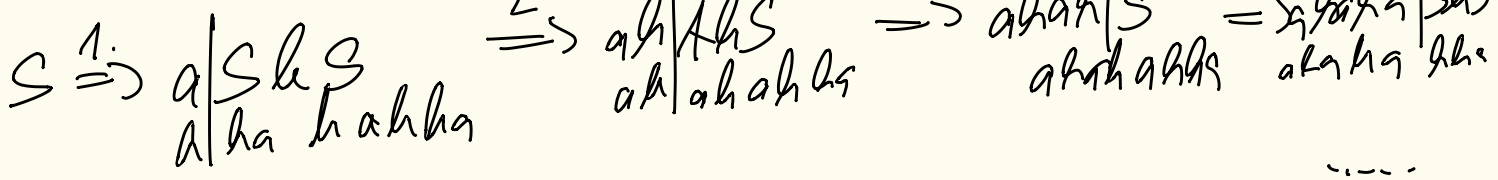
Očíslujme jednotlivá pravidla

1. $S \to aSbS$
2. $S \to bA$
3. $A \to bA$
4. $A \to a$



ahahahha

$$S \overset{1.}{\Rightarrow} \begin{array}{l|l} a & SbS \\ a & ha\ hahha \end{array} \overset{2}{\Rightarrow} \begin{array}{l|l} ah & AhS \\ ah & ahahha \end{array} \overset{4.}{\Rightarrow} \begin{array}{l|l} ahah & S \\ ahahahha & \end{array} \overset{1}{\Rightarrow} \begin{array}{l|l} ahaha & SbS \\ ahaha & hha \end{array}$$



.....

Analýza shora

Příklad. Je dána CF gramatika $\mathcal{G}$ s jedním neterminálem S, $\Sigma = \{a, b\}$ a pravidly:

$$S \to aSbS, S \to bA, A \to bA, A \to a.$$

Očíslujme jednotlivá pravidla

1. $S \to aSbS$
2. $S \to bA$
3. $A \to bA$
4. $A \to a$

Levý rozbor $w \in \Sigma^\star$ je posloupnost čísel pravidel použitých pro vygenerování slova w v levé derivaci.

Analýza shora

CF gramatika $\mathcal{G}$ je jednoduchá $LL(1)$ gramatika, jestliže pro každé její pravidlo platí:

1. Pravá strana začíná terminálním symbolem.
2. Jestliže dvě pravidla mají stejnou levou stranu, pak se liší terminálním symbolem, kterým začíná pravá strana pravidla.

Analýza shora

$LL(1)$ **analyzátor**

Dána CF gramatika $\mathcal{G} = (N, \Sigma, S, P)$. Pak $LL(1)$ **analyzátor vztahující se k** $\mathcal{G}$ je funkce $\mathcal{M}$, která každé dvojici (A, a), kde $A \in (N \cup \Sigma \cup \{\varepsilon\})$, $a \in (\Sigma \cup \{\varepsilon\})$ přiřadí

- krátit, chyba, přijmout nebo
- $(\beta; i)$, kde β je pravá strana i-tého pravidla

taková, že

- $\mathcal{M}(a, a) =$ krátit pro $a \in \Sigma$,
- $\mathcal{M}(\varepsilon, \varepsilon) =$ přijmout,
- if $\mathcal{M}(X, a) = (\beta; i)$, pak $X \to \beta$ je i-té pravidlo a β začíná a,
- ve všech ostatních případech je $\mathcal{M}(X, a) =$ chyba.

Analýza shora

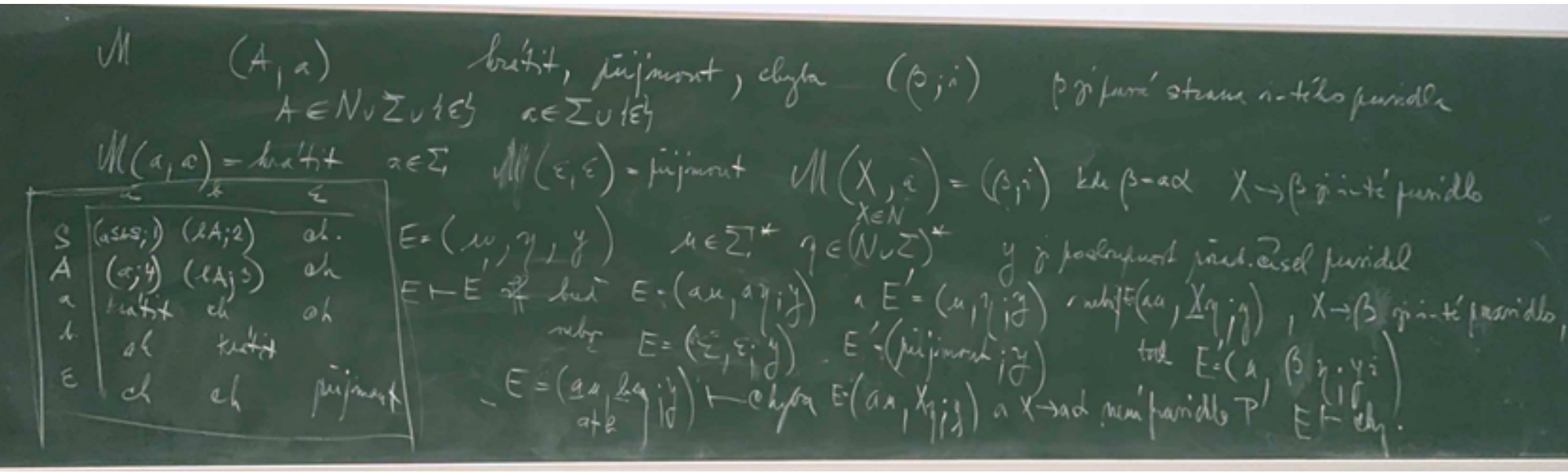
Pro gramatiku $\mathcal{G}$ z příkladu je $LL(1)$ analyzátor dán následující tabulkou:

	a	b	ε
S	$(aSbS; 1)$	$(bA; 2)$	chyba
A	$(a; 4)$	$(bA; 3)$	chyba
a	krátit	chyba	chyba
b	chyba	krátit	chyba
ε	chyba	chyba	přijmout

Analýza shora

Situace $LL(1)$ analyzátoru $\mathcal{M}$ je trojice

- (u, η, y), kde $u \in \Sigma^\star$, $\eta \in (N \cup \Sigma)^\star$ a y je posloupnost čísel pravidel
- (přijmout, y) nebo chyba.



Analýza shora

Situace $LL(1)$ analyzátoru $\mathcal{M}$ je trojice

- (u, η, y), kde $u \in \Sigma^\star$, $\eta \in (N \cup \Sigma)^\star$ a y je posloupnost čísel pravidel
- (přijmout, y) nebo chyba.

Jeden krok $\mathcal{M}$**:** E a E' jsou situace. Pak $E \vdash E'$, pokud

1. $E = (au, a\eta, y)$ pro $a \in \Sigma$ a $E' = (u, \eta, y)$,
2. $E = (au, X\eta, y)$, $X \in N$, $\mathcal{M}(X, a) = (\beta; i)$ a $E' = (u, \beta\eta, yi)$,
3. $E = (\varepsilon, \varepsilon, y)$ a $E' = (\text{přijmout}, y)$,
4. $E = (au, X\eta, y)$, $X \in N$, $\mathcal{M}(X, a) = \text{chyba}$ a E' je chyba.

Analýza shora

Počáteční situace $\mathcal{M}$ je (u, S, ε).

$\mathcal{M}$ je $LL(1)$ analyzátor pro $\mathcal{G}$, if

1. if $u \notin L(\mathcal{G})$, tak $(u, S, \varepsilon) \vdash^\star$ chyba,
2. if $u \in L(\mathcal{G})$, tak $(u, S, \varepsilon) \vdash^\star$ (přijmout, y), kde y je levý rozbor u v $\mathcal{G}$.

Analýza shora

Věta. Ke každé jednoduché $LL(1)$ gramatice existuje $LL(1)$ analyzátor, který je jejím analyzátorem.

Analýza shora

Definujme Fi(α) je množina všech terminálních symbolů *a* na něž začíná některé terminální slovo generované z α.
V případě, že z α se dá vygenerovat ε, tak $\varepsilon \in$ Fi(α).

Analýza shora

CF gramatika je *LL*(1) gramatika, jestliže:

1. Jestliže v $\mathcal{G}$ existují dvě pravidla $A \to \alpha$ a $A \to \beta$ se stejnou levou stranou, pak pro každé dvě levé derivace $S \Rightarrow^\star uA\mu$ a $S \Rightarrow^\star vA\chi$, $u, v \in \Sigma^\star$, $\mu, \chi \in (N \cup \Sigma)^\star$ platí $\mathrm{Fi}(\alpha\mu) \cap \mathrm{Fi}(\beta\chi) = \emptyset$.
2. Jestliže navíc $\mathrm{Fi}(A)$ obsahuje ε, pak $\mathrm{Fi}(\mu)$ je také disjunktní s $\mathrm{Fi}(\alpha\mu)$ (a tedy i $\mathrm{Fi}(\beta\chi)$).

Bezkontextové jazyky

Úlohy algoritmicky řešitelné pro bezkontextové gramatiky

- Generuje daná bezkontextová gramatika neprázdný jazyk?

Bezkontextové jazyky

Úlohy algoritmicky řešitelné pro bezkontextové gramatiky

- Generuje daná bezkontextová gramatika neprázdný jazyk?
- Platí $w \in L(\mathcal{G})$ pro danou bezkontextovou gramatiku $\mathcal{G}$ a slovo w?

Řada dalších úloh/problémů je algoritmicky neřešitelná.

$L = \{a^n b^n c^n \mid n \geq 1\} \notin CF$

$\uparrow$ kontextová $\quad S \to aSBC \mid aBC$

"$CB \to BC$"

$\cancel{B \to b \quad C \to c}$

$$\boxed{\begin{array}{ll} CB \to XB & XY \to BY \\ XB \to XY & BY \to BC \end{array}}$$

$aB \to ab \quad bC \to bc$

$bB \to bb \quad cC \to cc$

$\hookrightarrow$ završí $a^n b^n c^n$

$S \Rightarrow^* a^i S(BC)^i \cancel{\Rightarrow a^{i+1}(BC)^{i+1}}$

$aaaBCBCBC$

$\Rightarrow^* a^{i+1} B^{i+1} C^{i+1} \Rightarrow^*$

$\Rightarrow^* a^{i+1} b^{i+1} c^{i+1}$

$i \geq 0$

$a^i B^j C^k B$

$a^i b^j c^k B$