Akademia Górniczo-Hutnicza
im. Stanisława Staszica w Krakowie

Gramatyki LL(1)

Teoria kompilacji

Dr inż. Janusz Majewski
Katedra Informatyki

Nazwa gramatyki: LL(k)

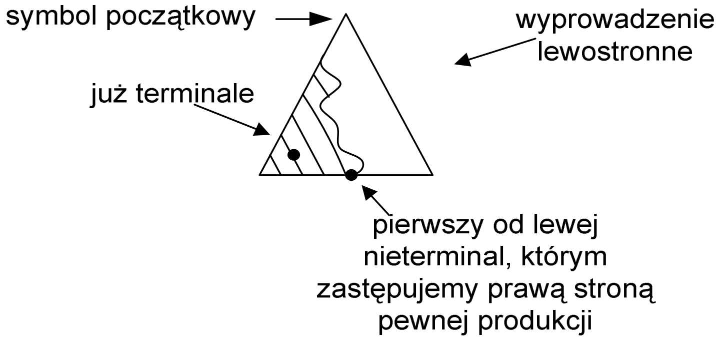
L L (k)

Przeglądanie wejścia od lewej strony do prawej

Odtwarzanie wywodu lewostronnego

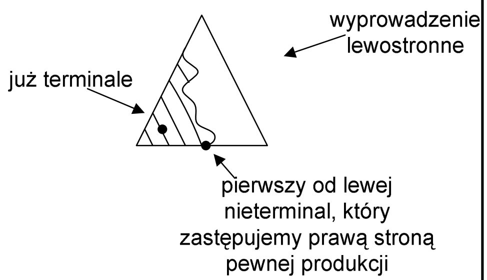
Wystarcza znajomosc "k" następnych symboli łańcucha wejściowego oraz skutków poprzednich kroków, aby wyznaczyć jednoznacznie produkcję, którą należy zastosować przy budowie drzewa wyprowadzenia

Zadanie analizy generacyjnej (zstępującej, top-down)



Odtworzenie wywodu lewostronnego metodą top-down

Istota wywodu top-down



Niech $G = < V, \Sigma, P, S > \in \mathcal{G}_{BK}$

$v = a_1 \ldots a_n$ – analizowane słowo

Załóżmy, że $v \in L(G)$.

Wówczas:

$$S = u_0 \Rightarrow_L^{p_1} u_1 \Rightarrow_L^{p_2} u_2 \Rightarrow_L \ldots \Rightarrow_L^{p_m} u_m = v$$

p_i – numery produkcji

$p_1 \ldots p_m$ – rozkład lewostronny słowa v

Poszukujemy tego rozkładu!

Istota definicji gramatyki LL(1)

Przypuśćmy, że: $u_i = a_1 \ldots a_j A\alpha$

(wyprowadzono już poprawnie pierwsze symbole)

Szukamy: u_{i+1} i p_{i+1} takich, że:

$$u_i \Rightarrow_L^{p_{i+1}} u_{i+1}$$

Znamy:

(1) $a_1 \ldots a_j$ – wyprowadzona, początkowa część słowa

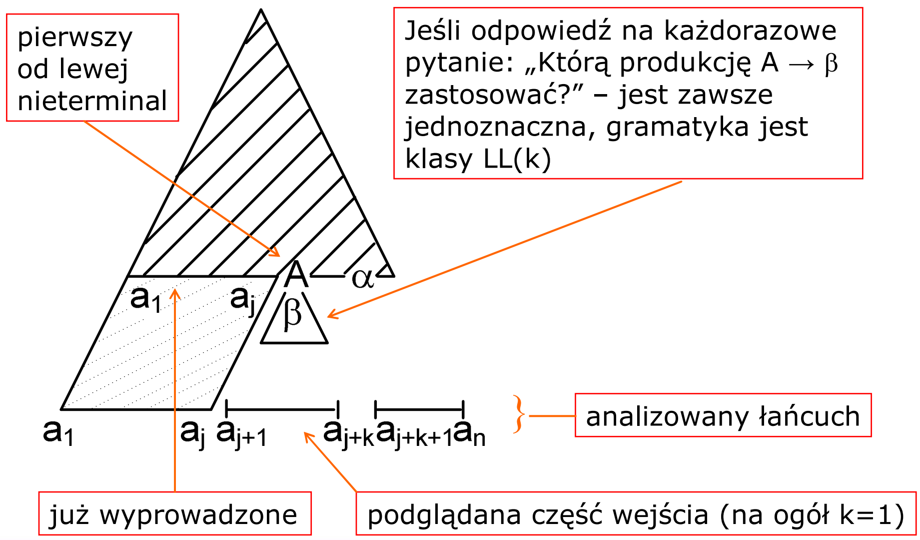
(2) $a_{j+1} \ldots a_{j+k}$ – następnych k symboli słowa wejściowego

(3) nieterminal A

Pytamy:

którą produkcję $A \rightarrow \beta$ należy zastosować w wyprowadzeniu. Jeśli odpowiedź na to pytanie jest <u>jednoznaczna</u> ($\forall$i) to gramatyka jest klasy LL(k)

Istota gramatyki i parsera LL(k)



Definicja gramatyki LL(1)

Niech:

$\omega, x, y \in \Sigma^*$

$\beta, \gamma \in (V \cup \Sigma)^*$

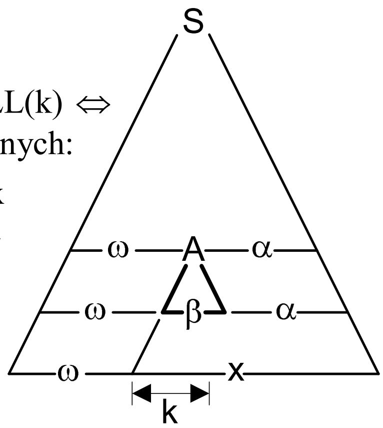
$A \in V$

$G = < V, \Sigma, P, S > \in \mathcal{G}_{BK}$ jest klasy LL(k) $\Leftrightarrow$ gdy dla każdej pary wywodów lewostronnych:

(1) $S \Rightarrow_L^* \omega A \alpha \Rightarrow_L \omega\beta\alpha \Rightarrow_L^* \omega x$

(2) $S \Rightarrow_L^* \omega A \alpha \Rightarrow_L \omega\gamma\alpha \Rightarrow_L^* \omega y$

jeżeli $FIRST_k(x) = FIRST_k(y)$ to $\beta = \gamma$



Przykład

$G = < \{ S, D\}, \{ a, b\}, P, S >$

(1) $S \rightarrow a\ D\ S$
(2) $S \rightarrow b$
(3) $D \rightarrow a$
(4) $D \rightarrow b\ S\ D$

Pytanie: czy G jest klasy LL(1) ?

$$\underbrace{S}_{s} \Rightarrow_L^* \underbrace{a}_{\omega}\ \underbrace{D}_{A}\ \underbrace{S}_{\alpha} \begin{cases} \Rightarrow_L^{(3)} \underbrace{a}_{\omega}\ \underbrace{a}_{\beta}\ \underbrace{S}_{\alpha} \Rightarrow_L^* \underbrace{a}_{\omega}\underbrace{\boxed{a}b}_{x} \\ \Rightarrow_L^{(3)} \underbrace{a}_{\omega}\ \underbrace{a}_{\gamma}\ \underbrace{S}_{\alpha} \Rightarrow_L^* \underbrace{a}_{\omega}\underbrace{\boxed{a}a\ a\ b}_{y} \\ \Rightarrow_L^{(4)} \underbrace{a}_{\omega}\ \underbrace{bSD}_{\beta}\ \underbrace{S}_{\alpha} \Rightarrow_L^* \underbrace{a}_{\omega}\underbrace{\boxed{b}b\ a\ b}_{x} \\ \Rightarrow_L^{(4)} \underbrace{a}_{\omega}\ \underbrace{bSD}_{\gamma}\ \underbrace{S}_{\alpha} \Rightarrow_L^* \underbrace{a}_{\omega}\underbrace{\boxed{b}a\ a\ b\ a\ b}_{y} \end{cases}$$

Jeżeli FIRST(x) = FIRST(y) = a to produkcja (3)

Jeżeli FIRST(x) = FIRST(y) = b to produkcja (4)

Takie rozumowanie można uogólnić na wszystkie możliwe wywody

$S \Rightarrow_L^* \omega S \alpha \Rightarrow \ldots$ oraz $S \Rightarrow_L^* \omega D \alpha \Rightarrow \ldots$

Więc gramatyka jest LL(1) !

Przykład

$G = < \{ S, D, E \}, \{ 0, 1, a, b \}, P, S >$

(1) $S \to D$
(2) $S \to E$
(3) $D \to a\,D\,b$
(4) $D \to 0$
(5) $E \to a\,E\,b\,b$
(6) $E \to 1$

$$\underset{S}{S} \Rightarrow_L^{*} \underset{\omega A \alpha}{S} \begin{cases} \Rightarrow_L^{(1)} \underset{\omega\,\beta\,\alpha}{D} \Rightarrow_L^{*} \underset{\omega\ \ x}{\boxed{a^k} 0\, b^k} \\ \Rightarrow_L^{(2)} \underset{\omega\,\gamma\,\alpha}{E} \Rightarrow_L^{*} \underset{\omega\ \ y}{\boxed{a^k} 1\, b^{2k}} \end{cases}$$

$L(G) = \{ a^n\, 0\, b^n : n \geq 0 \} \cup \{ a^n\, 1\, b^{2n} : n \geq 0 \}$
$(\omega = \varepsilon ; \alpha = \varepsilon ;)$
Dla dowolnego k $FIRST_k(x) = FIRST_k(y) = a^k$, ale $\beta \neq \gamma$ $(D \neq E)$.
Ponieważ k było dowolne – gramatyka G <u>nie</u> jest LL(k) dla żadnego k !

Twierdzenie 1

Twierdzenie pozwalające stwierdzić, czy gramatyka G jest klasy LL(k) dla danego k

$G \in \mathcal{G}_{BK}$ jest klasy LL(k) $\Leftrightarrow$

$\forall (A \to \beta) \in P, (A \to \gamma) \in P: \ \beta \neq \gamma$

$\forall\ \alpha : S \Rightarrow_L^{*} \omega A \alpha, \quad \omega \in \Sigma^{*}$

zachodzi: $FIRST_k(\beta\alpha) \cap FIRST_k(\gamma\alpha) = \varnothing$

Przykład

(1) $S \rightarrow S\ a$
(2) $S \rightarrow b$

Rozważamy: wyprowadzenie produkcje

$S \Rightarrow_L^i \ S\ a^i$	$S \rightarrow S\ a$	$S \rightarrow b$
$S \Rightarrow_L^* \ \omega A \alpha$	$A \rightarrow \beta$	$A \rightarrow \gamma$

Dla $k \leq i$ mamy

$FIRST_k(\beta\alpha) = FIRST_k(Saa^i) = \{ba^{k-1}\}$

$FIRST_k(\gamma\alpha) = FIRST_k(ba^i) = \{ba^{k-1}\}$

Ponieważ i – dowolne, $k \leq i$, więc G nie może być klasy LL(k) dla żadnego k.

Uwaga: żadna gramatyka z lewostronną rekursją nie jest klasy LL(k) dla żadnego k.

Twierdzenie 2

$G \in \mathcal{G}_{BK}$ jest klasy LL(1) $\Leftrightarrow$
$\forall\ (A \rightarrow \alpha_1 \mid \alpha_2 \mid \ldots \mid \alpha_n) \in P$ zachodzi:

1) $FIRST_1(\alpha_1), FIRST_1(\alpha_2), \ldots, FIRST_1(\alpha_n)$ są parami rozłączne, oraz
2) jeżeli $\alpha_i \Rightarrow^* \varepsilon$ to
$FIRST_1(\alpha_j) \cap FOLLOW_1(A) = \varnothing$ (pusty)
dla $1 \leq j \leq n \ \wedge \ i \neq j$

Przykład

Poprzednia gramatyka

$S \to S\ a \mid b$

po przekształceniu ma postać:

$S \to b\ S'$ $\quad\quad \underset{A}{S'} \to \underset{\alpha_1}{a\ S'} \mid \underset{\alpha_2}{\varepsilon}$

i jest klasy LL(1) bo:

(1) $FIRST_1(\alpha_1) = FIRST_1(aS') = \{a\}$

$FIRST_1(\alpha_2) = FIRST_1(\varepsilon) = \{\varepsilon\}$

$FIRST_1(\alpha_1) \cap FIRST_1(\alpha_2) = \varnothing$

(2) $\varepsilon \Rightarrow^* \varepsilon$

$FIRST_1(\alpha_1) = FIRST_1(aS') = \{a\}$

$FOLLOW_1(A) = FOLLOW_1(S') = \{\$\}$

$FIRST_1(\alpha_1) \cap FOLLOW_1(A) = FIRST_1(aS') \cap FOLLOW_1(S') = \varnothing$

Twierdzenie 3

(1) $G \in \mathcal{G}_{BK}$ jest klasy LL(k) $\Rightarrow \exists\ G' \in \mathcal{G}_{BK}$:

(a) G' - jest klasy LL(k+1)

(b) G' - nie zawiera ε-produkcji

(c) $L(G') = L(G)$

(2) $G \in \mathcal{G}_{BK}$ jest klasy LL(k), gdzie $k \geq 2$,

oraz G nie zawiera ε-produkcji $\Rightarrow \exists\ G' \in \mathcal{G}_{BK}$:

(a) G' - jest klasy LL(k-1)

(b) $L(G') = L(G)$

Lewostronna faktoryzacja

Przekształceniem obniżającym stopień gramatyki LL kosztem wprowadzenia ε-produkcji jest lewostronna faktoryzacja. Polega ona na zamianie produkcji postaci:

$$A \to \alpha\beta_1 \mid \ldots \mid \alpha\beta_k$$

na produkcje:

$$A \to \alpha A'$$

$$A' \to \beta_1 \mid \ldots \mid \beta_k$$

Przykład

Gramatyka:

$$S \to a\,S \mid a$$

$$A \to \alpha\,\beta_1 \mid \alpha\,\beta_2$$

jest gramatyką klasy LL(2) bez ε-produkcji.

Po lewostronnej faktoryzacji

$$S \to a\,S' \qquad S' \to S \mid \varepsilon$$

$$A \to \alpha\,A' \qquad A' \to \beta_1 \mid \beta_2$$

jest klasy LL(1), ale zawiera ε-produkcje