



AKADEMIA GÓRNICZO-HUTNICZA
IM. STANISŁAWA STASZICA W KRAKOWIE

Parsery LALR(1)

Teoria kompilacji

Dr inż. Janusz Majewski
Katedra Informatyki



Gramatyki LALR(k) (Look Ahead LR(k))

Duża klasa gramatyk mających praktyczne zastosowanie nie należy do SLR(1), ale dla tych gramatyk daje się skonstruować parser o rozmiarze tablicy sterującej (funkcja "f" i "g") identycznym z SLR(1), tyle że przy pomocy trochę bardziej skomplikowanego algorytmu. Klasa ta nazywa się LALR(1) i z praktycznego punktu widzenia jest najważniejszą podklasą LR(1), gdyż:

- 1) Jest dostatecznie szeroka, żeby objąć znaczną większość języków programowania.
- 2) Rozmiar tablicy sterującej parsera jest jeszcze do przyjęcia (w odróżnieniu od kanonicznego LR(1)).

Jądro zbioru LR(1)-sytuacji

CORE – jądro zbioru LR(1) – sytuacji

A - zbiór LR(1) – sytuacji

$$CORE(A) = \{ [A \rightarrow \alpha \bullet \beta] \mid [A \rightarrow \alpha \bullet \beta, u] \in A \}$$

Twierdzenie

J_0 – kanoniczny system zbiorów LR(0) – sytuacji w G

J_1 – kanoniczny system zbiorów LR(1) – sytuacji w G

$$J_0 = \{ CORE(A) : A \in J_1 \}$$

CORE(GOTO(...))

Twierdzenie

$$CORE(GOTO_1(A, x)) = GOTO_0(CORE(A), x) \\ x \in (V \cup \Sigma)$$

gdzie:

$GOTO_1$ – funkcja $GOTO$ dla zbioru LR(1)-sytuacji

$GOTO_0$ – funkcja $GOTO$ dla zbioru LR(0)-sytuacji

e – relacja równości jąder

$$e \subset J_1 \times J_1 : A_1 e A_2 \stackrel{!}{\Leftrightarrow} CORE(A_1) = CORE(A_2)$$

Gramatyka LALR(1)

$$G = \langle V, \Sigma, P, S \rangle \in \mathbf{G}_{BK}$$

J_1 – kanoniczny system zbiorów LR(1) – sytuacji dla G

$$J = \{ B([A]_e) : A \in J_1 \};$$

gdzie:

$$B([A]_e) = \bigcup_{C \in [A]_e} (C \in J_1)$$

G – gramatyka LALR(1) $\Leftrightarrow (\forall B \in J) (B - \text{zgodny})$



Przykład (łączenie zbiorów LR(1)-sytuacji)

$$J_1 = \{ A_1, A_2, A_3, A_4, A_5, A_6 \}$$

$$e = \{ (A_1, A_1), (A_2, A_2), (A_3, A_3), (A_4, A_4), (A_2, A_3), \\ (A_3, A_2), (A_2, A_4), (A_4, A_2), (A_3, A_4), (A_4, A_3), \\ (A_5, A_5), (A_6, A_5), (A_5, A_6), (A_6, A_6) \}$$

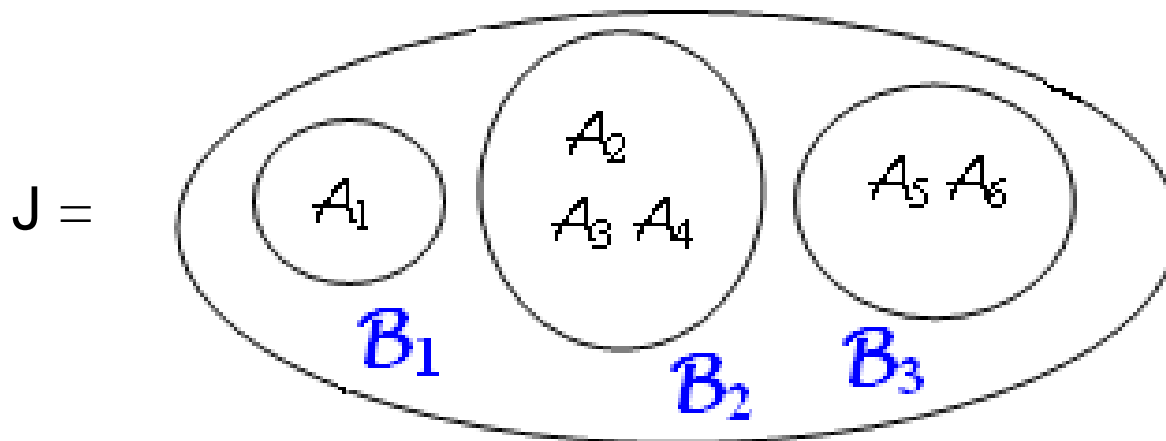
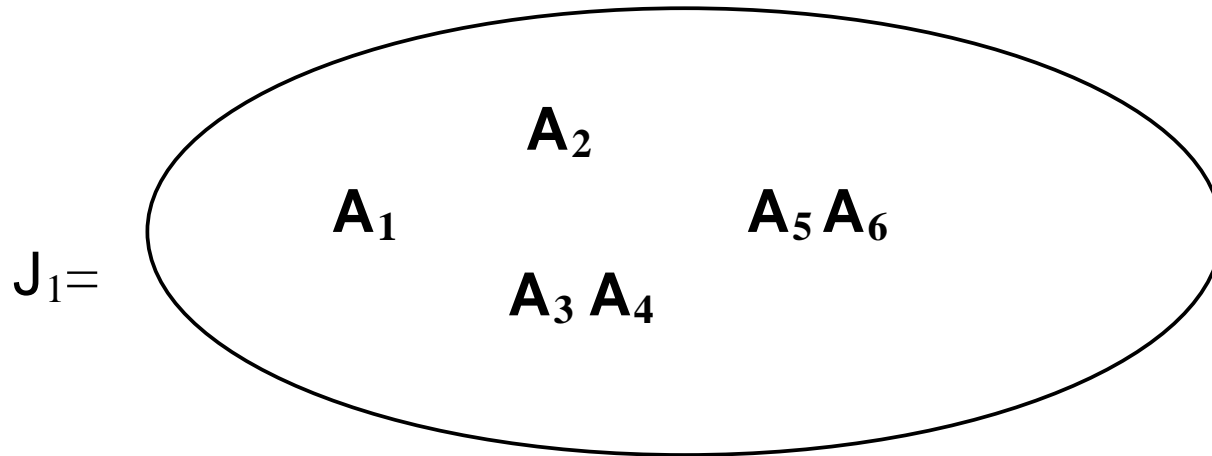
$$B_1 = A_1$$

$$B_2 = A_2 \cup A_3 \cup A_4$$

$$B_3 = A_5 \cup A_6$$

$$J = \{ B_1, B_2, B_3 \}$$

Przykład c.d.



Konstrukcja tablicy parsera LALR(1)

WEJŚCIE: $G = \langle V, T, P, S \rangle \in \mathbf{G}_{BK}$

WYJŚCIE: Tablica LALR(1)

ALGORYTM:

1. Wyznaczamy J_1 – kanoniczny system zbiorów LR(1)-sytuacji dopuszczalnych
2. Konstruujemy zbiór klas abstrakcji dla relacji e :

$$J := \{ B([A]_e) : A \in J_1 \};$$

$$/* J = \{ B_0, B_1, \dots, B_n \}$$

$$\text{gdzie } B_i = A_{(i1)} \cup A_{(i2)} \cup \dots \cup A_{(i m_i)}$$

$$i = 0, 1, \dots, n$$

$$[A_{(i1)}]_e = [A_{(i2)}]_e = \dots = [A_{(i m_i)}]_e */$$

3. Określamy funkcje f i g dla $J = \{ B_0, B_1, \dots, B_n \}$ identycznie, jak to miało miejsce w przypadku kanonicznego LR(1);

Przykład

$$G' = \langle V' = \{ S', S, L, R \}, T = \{ =, *, \underline{id} \}, P', S' \rangle$$

$$P' = \{$$

$$(0) \quad S' \rightarrow S$$

$$(1) \quad S \rightarrow L = R$$

$$(2) \quad S \rightarrow R$$

$$(3) \quad L \rightarrow *R$$

$$(4) \quad L \rightarrow \underline{id}$$

$$(5) \quad R \rightarrow L$$

$$\}$$

$$FIRST_1(S) = \{ *, \underline{id} \}$$

$$FIRST_1(L) = \{ *, \underline{id} \}$$

$$FIRST_1(R) = \{ *, \underline{id} \}$$

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

$$FOLLOW_1(L) = \{ \$, = \}$$

$$FOLLOW_1(R) = \{ \$, = \}$$

Przykład

Sprawdzenie czy poniższa gramatyka G' jest SLR(1)?

Próbujemy skonstruować J_0 – kanoniczny system LR(0) – sytuacji

$$A_0 = \{ \begin{array}{l} [S' \rightarrow \bullet S], \\ [S \rightarrow \bullet L = R], \\ [S \rightarrow \bullet R], \\ [L \rightarrow \bullet *R], \\ [L \rightarrow \bullet \underline{id}], \\ [R \rightarrow \bullet L] \end{array} \}$$

$$A_1 = GOTO(A_0, S) = \{ [S' \rightarrow S \bullet] \}$$

$$A_2 = GOTO(A_0, L) = \{ [S \rightarrow L \bullet = R], \\ [R \rightarrow L \bullet] \}$$

- 1) $S' \rightarrow S$
- 2) $S \rightarrow L = R$
- 3) $S \rightarrow R$
- 4) $L \rightarrow *R$
- 5) $L \rightarrow \underline{id}$
- 6) $R \rightarrow L$

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

$$FOLLOW_1(L) = \{ \$, = \}$$

$$FOLLOW_1(R) = \{ \$, = \}$$

Przykład

$$A_2 = GOTO(A_0, L) = \{ [S \rightarrow L \bullet = R] , \\ [R \rightarrow L \bullet] \\ \}$$

Ponieważ:

$$FOLLOW_1(R) = \{ \$, = \}$$

Więc:

$$\left. \begin{array}{l} f(T_2, =) = \underline{red-6} \\ f(T_2, =) = \underline{shift} \end{array} \right\} \underline{\text{konflikt!}}$$

gramatyka nie jest SLR(1)!

Przykład

Konstruujemy J_1 – kanoniczny system zbiorów LR(1)–sytuacji.

$$A_0 = \{ [S' \rightarrow \bullet S, \$], \\ [S \rightarrow \bullet L=R, \$], \\ [S \rightarrow \bullet R, \$], \\ [L \rightarrow \bullet *R, \$/=], \\ [L \rightarrow \bullet \underline{id}, \$/=], \\ [R \rightarrow \bullet L, \$] \}$$

$$A_1 = \{ [S' \rightarrow S \bullet, \$] \}$$

$$A_2 = \{ [S \rightarrow L \bullet =R, \$] \\ [R \rightarrow L \bullet, \$] \}$$

$$A_1 = GOTO(A_0, S)$$

$$A_2 = GOTO(A_0, L)$$

- 1) $S' \rightarrow S$
- 2) $S \rightarrow L = R$
- 3) $S \rightarrow R$
- 4) $L \rightarrow *R$
- 5) $L \rightarrow \underline{id}$
- 6) $R \rightarrow L$

Przykład

$$A_3 = \{ [S \rightarrow R \bullet, \$] \}$$

$$A_3 = GOTO(A_0, R)$$

$$A_4 = \{ [L \rightarrow * \bullet R, \$/ =], \\ [R \rightarrow \bullet L, \$/ =], \\ [L \rightarrow \bullet * R, \$/ =], \\ [L \rightarrow \bullet \underline{id}, \$/ =] \}$$

$$A_4 = GOTO(A_0, *)$$

$$A_5 = \{ [L \rightarrow \underline{id} \bullet, \$/ =] \}$$

$$A_5 = GOTO(A_0, \underline{id})$$

$$A_6 = \{ [S \rightarrow L = \bullet R, \$], \\ [R \rightarrow \bullet L, \$], \\ [L \rightarrow \bullet * R, \$], \\ [L \rightarrow \bullet \underline{id}, \$] \}$$

$$A_6 = GOTO(A_2, =)$$

- 1) $S' \rightarrow S$
- 2) $S \rightarrow L = R$
- 3) $S \rightarrow R$
- 4) $L \rightarrow *R$
- 5) $L \rightarrow \underline{id}$
- 6) $R \rightarrow L$

Przykład

$$A_7 = \{ [L \rightarrow *R\bullet, \$| =] \}$$

$$A_8 = \{ [R \rightarrow L\bullet, \$| =] \}$$

$$A_9 = \{ [S \rightarrow L=R\bullet, \$] \}$$

$$A_{10} = \{ [L \rightarrow *\bullet R, \$],$$

$$[R \rightarrow \bullet L, \$],$$

$$[L \rightarrow \bullet *R, \$],$$

$$[L \rightarrow \bullet \underline{id}, \$] \}$$

$$A_7 = GOTO(A_4, R)$$

$$A_8 = GOTO(A_4, L)$$

$$A_4 = GOTO(A_4, *)$$

$$A_5 = GOTO(A_4, \underline{id})$$

$$A_9 = GOTO(A_6, R)$$

$$A_{10} = GOTO(A_6, *)$$

- 1) $S' \rightarrow S$
- 2) $S \rightarrow L = R$
- 3) $S \rightarrow R$
- 4) $L \rightarrow *R$
- 5) $L \rightarrow \underline{id}$
- 6) $R \rightarrow L$

Przykład

$$A_{11} = \{[L \rightarrow \underline{id}\bullet, \$]\}$$

$$A_{12} = \{[R \rightarrow L\bullet, \$]\}$$

$$A_{13} = \{[L \rightarrow *R\bullet, \$]\}$$

$$A_{11} = GOTO(A_6, \underline{id})$$

$$A_{12} = GOTO(A_6, L)$$

$$A_{13} = GOTO(A_{10}, R)$$

$$A_{12} = GOTO(A_{10}, L)$$

$$A_{10} = GOTO(A_{10}, *)$$

$$A_{11} = GOTO(A_{10}, \underline{id})$$

- 1) $S' \rightarrow S$
- 2) $S \rightarrow L = R$
- 3) $S \rightarrow R$
- 4) $L \rightarrow *R$
- 5) $L \rightarrow \underline{id}$
- 6) $R \rightarrow L$

Przykład

$$B_0 = A_0$$

$$B_1 = A_1$$

$$B_2 = A_2$$

$$B_3 = A_3$$

$$\underline{B_4 = A_4 \cup A_{10}}$$

$$\underline{B_5 = A_5 \cup A_{11}}$$

$$B_6 = A_6$$

$$\underline{B_7 = A_7 \cup A_{13}}$$

$$\underline{B_8 = A_8 \cup A_{12}}$$

$$B_9 = A_9$$

$$B_1 = \text{GOTO}(B_0, S)$$

$$B_2 = \text{GOTO}(B_0, L)$$

$$B_3 = \text{GOTO}(B_0, R)$$

$$B_4 = \text{GOTO}(B_0, *)$$

$$B_5 = \text{GOTO}(B_0, \underline{\text{id}})$$

$$B_6 = \text{GOTO}(B_2, =)$$

$$B_7 = \text{GOTO}(B_4, R)$$

$$B_8 = \text{GOTO}(B_4, L)$$

$$B_4 = \text{GOTO}(B_4, *)$$

$$B_5 = \text{GOTO}(B_4, \underline{\text{id}})$$

$$B_9 = \text{GOTO}(B_6, R)$$

$$B_4 = \text{GOTO}(B_6, *)$$

$$B_5 = \text{GOTO}(B_6, \underline{\text{id}})$$

$$B_8 = \text{GOTO}(B_6, L)$$

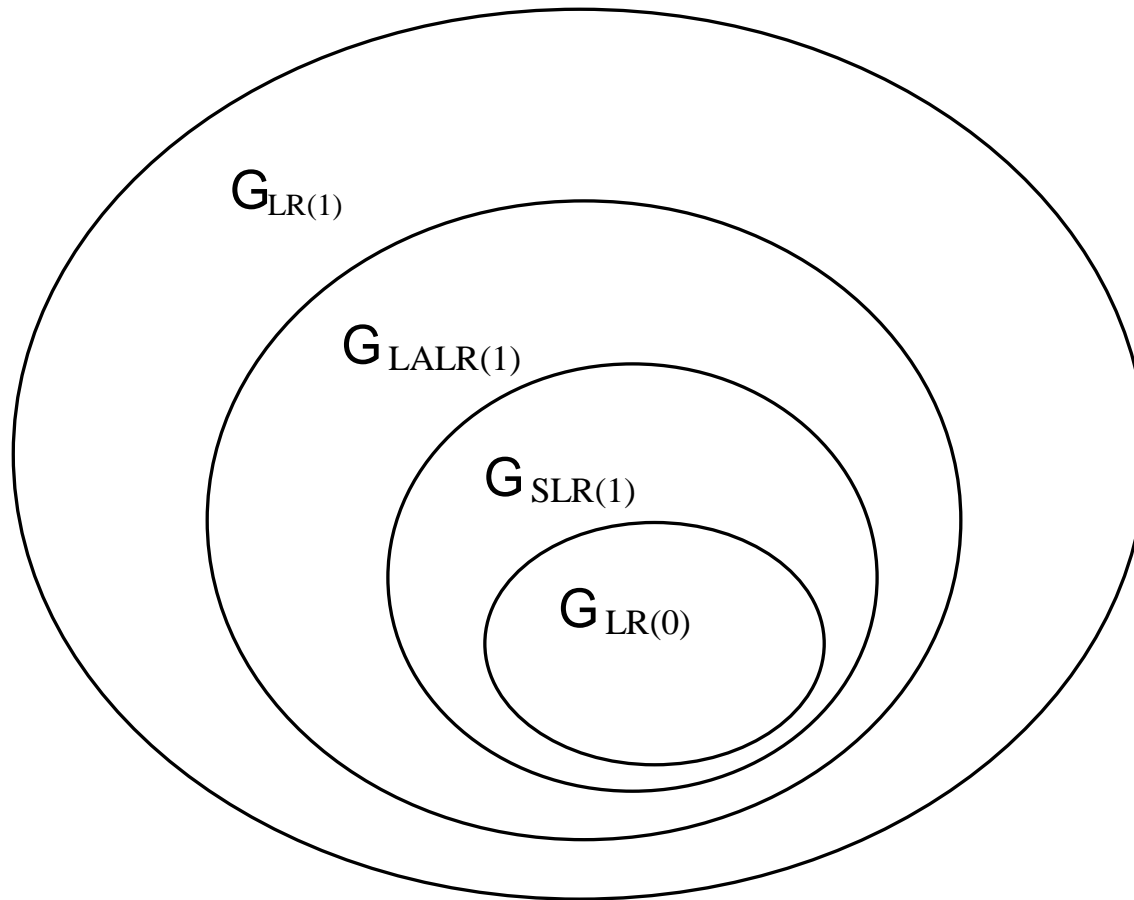
Przykład

TABLICA LALR(1)

stan	f				g		
	\$	=	*	<u>id</u>	S	L	R
T ₀			<u>shift-4</u>	<u>shift-5</u>	T ₁	T ₂	T ₃
T ₁	<u>acc</u>						
T ₂	<u>red-5</u>	<u>shift-6</u>					
T ₃	<u>red-2</u>						
T ₄			<u>shift-4</u>	<u>shift-5</u>		T ₈	T ₇
T ₅	<u>red-4</u>	<u>red-4</u>					
T ₆			<u>shift-4</u>	<u>shift-5</u>		T ₈	T ₉
T ₇	<u>red-3</u>	<u>red-3</u>					
T ₈	<u>red-5</u>	<u>red-5</u>					
T ₉	<u>red-1</u>						

(Kanoniczny LR(1) ma o 4 stany więcej)

Zależności pomiędzy klasami gramatyk



Każda gramatyka LR(0)
jest gramatyką SLR(1).
Każda gramatyka SLR(1)
jest gramatyką LALR(1).
Każda gramatyka LALR(1)
jest gramatyką LR(1).

$$G_{LR(0)} \subset G_{SLR(1)} \subset G_{LALR(1)} \subset G_{LR(1)}$$

$$G_{LR(0)} \neq G_{SLR(1)} \neq G_{LALR(1)} \neq G_{LR(1)}$$



Języki bezkontekstowe zdeterminowane a języki LR

Język bezkontekstowy nazywamy zdeterminowanym, gdy jest on akceptowany przez deterministyczny automat ze stosem.

$$L_{\text{BK DET}} = L_{\text{LR}(1)}$$

Każdy język bezkontekstowy zdeterminowany ma swoją gramatykę LR(1).

Każdy język bezkontekstowy zdeterminowany posiadający własność przedrostkową ma swoją gramatykę LR(0).

Języki LL a języki LR

$$L_{LL(K)} \subset L_{LR(K)}$$

$$L_{LL(K)} \neq L_{LR(K)}$$

Każda gramatyka LL(k) jest gramatyką LR(k).
Istnieją języki LR(k), które nie są generowane przez żadną gramatykę LL(k).

Języki LL(0) są jednoelementowe.

Podklasy języków bezkontekstowych

