

LABORATOR NR 2

ANALIZA SINTACTICĂ DESCENDENTĂ

Analiza sintactică.

Fie $G = (N, \Sigma, P, S)$ o gramatică independentă de context și $\alpha \in L(G)$. Spunem că a fost efectuată analiza sintactică a secvenței α dacă a fost găsit cel puțin un arbore generator pentru secvența dată. Se lucrează cu analiza sintactică la stânga (arborele generator se construiește de la rădăcină spre nodurile frunză) și analiză sintactică la dreapta (arborele generator se construiește de la nodurile frunză spre rădăcină).

Presupunem că G are producțiile numerotate de la 1 la p și $\alpha, \beta \in (N \cup \Sigma)^*$.

Scriem $\alpha^i \Rightarrow \beta$ dacă se folosește producția cu numărul i , iar neterminalul care derivează este cel mai din stânga. Analog pentru derivarea la dreapta.

Dacă $\pi = i_1 i_2 \dots i_n$, notăm $\tilde{\pi} = i_n i_{n-1} \dots i_1$. Pentru $\alpha \in L(G)$, se numește analiză sintactică la stânga o secvență $\pi \in \{1, 2, \dots, p\}^*$ astfel încât $S \xRightarrow{\pi} \alpha$. Se numește analiză sintactică la dreapta a cuvântului α o secvență $\pi \in \{1, 2, \dots, p\}^*$ astfel încât $S \xRightarrow{\pi} \alpha$.

1. Analiza sintactică descendentă fără reveniri.

Teoremă. Fie $G = (N, \Sigma, P, S)$ o gramatică independentă de context cu producțiile numerotate de la 1 la p . Definim SSTOS $\delta = (N, \Sigma, \{1, 2, \dots, p\}, R, S)$ unde R constă din regulile $A \rightarrow \alpha, \beta$ astfel încât dacă $A \rightarrow \alpha$ este producția i , atunci $\beta = i\alpha'$ unde $\beta = i\alpha'$ unde α' se obține din α prin ștergerea tuturor terminalelor. Atunci $\tau(S) = \{(w, \pi) \mid S \xRightarrow{\pi} w\}$

Aplicație. Fie gramatica expresiilor aritmetice simple:

1. $E \rightarrow E + T$
2. $E \rightarrow T$
3. $T \rightarrow T * F$
4. $T \rightarrow F$
5. $F \rightarrow (E)$
6. $F \rightarrow a$

Se construiește schema $S = \{E, T, F\}, \{+, *, (,), a\}; \{1, 2, \dots, 6\}; R; E$ unde :

1. $E \rightarrow E + T, 1ET$
2. $E \rightarrow T, 2T$
3. $T \rightarrow T * F, 3TF$
4. $T \rightarrow F, 4F$
5. $F \rightarrow (E), 5E$
6. $F \rightarrow a, 6$

Pentru secvența $w = a + a$ se obține traducerea:

$$(E, E) \Rightarrow (E + T, 1ET) \Rightarrow (T^1 + T, 12T^1T) \Rightarrow (F + T, 124FT) \Rightarrow (a + T, 1246T) \Rightarrow \\ \Rightarrow (a + F, 1246F) \Rightarrow (a + a, 124646)$$

Traducerea secvenței w este $\tau(w) = 124646$.

Teoremă. Fie $G = (N, \Sigma, P, S)$ o gramatică independentă de context cu producțiile numerotate de la 1 la p . Definim traducătorul pushdown nedeterminist $\rho = (\{q\}; \Sigma; N \cup \Sigma; \{1, 2, \dots, p\}; \delta; q; S; \phi)$ unde:

1. $(q, \alpha, i) \in \delta(q, \varepsilon, A) \quad i : (A \rightarrow \alpha) \in P$
2. $(q, \varepsilon, \varepsilon) \in \delta(q, a, a) \quad \forall a \in \Sigma$

Atunci $\tau_\varepsilon(\rho) = \{(w, \pi) \mid S \xRightarrow{\pi} w\}$.

Aplicație. Fie gramatica expresiilor aritmetice simple:

1. $E \rightarrow E + T$
2. $E \rightarrow T$
3. $T \rightarrow T * F$
4. $T \rightarrow F$
5. $F \rightarrow (E)$
6. $F \rightarrow a$

Se construiește traducătorul pushdown nedeterminist:

$$P = (\{q\}; \{+, *, (,), a\}; \{E, T, F, +, *, (,), a\}; \{1, 2, \dots, 6\}; \delta; q; E; \phi)$$

1. $\delta(q, \varepsilon, E) = \{(q, E + T, 1); (q, T, 2)\}$
2. $\delta(q, \varepsilon, T) = \{(q, T * F, 3); (q, F, 4)\}$
3. $\delta(q, \varepsilon, F) = \{(q, (E), 5); (q, a, 6)\}$
4. $\delta(q, x, x) = \{(q, \varepsilon, \varepsilon)\} \forall x \in \Sigma$

Pentru secvența $w = a + a$ se obține traducerea 124646 astfel:

$$(q, a + a, E, \varepsilon) \mid - (q, a + a, E + T, 1) \mid - (q, a + a, T + T, 12) \mid - (q, a + a, F + T, 124) \mid - (q, a + a, a + T, 1246) \mid - (q, a, T, 1246) \mid - (q, a, F, 12464) \mid - (q, a, a, 124646) \mid - (q, \varepsilon, \varepsilon, 124646)$$

Algoritmul de analiză sintactică descendentă cu reveniri

Intrare: G o gramatică independentă de context, fără recursie stângă și un cuvânt $w = a_1 a_2 \dots a_n$, $n \geq 0$. Presupunem că producțiile sunt numerotate de la 1 până la p.

Ieșire: analiza sintactică la stânga a lui w dacă aceasta există, sau “eroare” în caz contrar.

Metoda: Pentru fiecare neterminal a, dacă $A \rightarrow \alpha_1 \mid \alpha_2 \mid \dots \mid \alpha_k$ sunt toate A-producțiile, se notează cu A_i alternativa $A \rightarrow \alpha_i$. Algoritmul utilizează configurații de forma (s, i, α, β) unde:

- s este starea algoritmului și $s \in \{q, b, t\}$, q-mers normal, b-mers înapoi, t-terminare

- $i \in \{1, 2, \dots, n+1\}$ este poziția curentă pe banda de intrare. Al $n+1$ -lea simbol de intrare este \$ folosit pe post de delimitator la dreapta

- $\alpha \in (\Sigma \cup \{A_i \mid \text{este varianta cu nr. } i \text{ asociata neterminal ulu } i A\})^*$ reprezintă conținutul benzii pushdown L_1 , bandă cu vârful spre dreapta, ce înregistrează alternativele încercate și simbolurile de intrare verificate

- $\beta \in (N \cup \Sigma \cup \{\$\})^*$ reprezintă conținutul unei benzi pushdown L_2 , cu vârful spre stânga. Simbolul din vârf reprezintă nodul activ al arborelui de derivare.

$(q, 1, \varepsilon, S\$)$ - configurația inițială

Trecerea de la o configurație la alta se realizează conform următorilor pași.

(P1): expandare

$$(q, i, \alpha, A\beta) \vdash (q, i, \alpha A_1, \gamma_1 \beta)$$

unde $(A \rightarrow \gamma_1) \in P$ este prima alternativă pentru A

(P2) concordanța între simbolul de intrare și cel derivat

$$(q, i, \alpha, a\beta) \vdash (q, i+1, \alpha a, \beta)$$

Se efectuează când $a_i = a, 1 \leq i \leq n$ și constă în mutarea simbolului terminal din vârful benzii L_2 în vârful benzii L_1 , și mutarea indicatorului pe banda de intrare.

(P3) obținerea unei configurații de acceptare

$$(q, n+1, \alpha, \$) \vdash (t, n+1, \alpha, \varepsilon)$$

Se aplică homomorfismul h lui α :

$h(a) = \varepsilon$ pentru $\forall a \in \Sigma$, $h(A_i) = p$ dacă p este producția $A_i : (A \rightarrow \gamma)$

(P4) Simbolul de intrare nu coincide cu cel derivat

$$(q, i, \alpha a, \beta) \vdash (q, i+1, \alpha, a\beta) \text{ pentru } a_i \neq a$$

(P5) Întoarcerea pe banda de intrare cu o poziție

$$(q, i, \alpha a, \beta) \vdash (b, i-1, \alpha, a\beta) \forall a \in \Sigma$$

(P6) Încercarea alternativei următoare

$$(b, i, \alpha A_j, \gamma_j \beta) \vdash (q, i, \alpha A_{j+1}, \gamma_{j+1} \beta)$$

dacă există a j+1 – a alternativă pentru A.

$$(b, i, \alpha A_j, \gamma_j \beta) \vdash \text{fără continuare}$$

dacă $i = 1$, $A = S$ și există numai j alternative pentru A. Se semnalează eroare.

$$(b, i, \alpha A_j, \gamma_j \beta) \vdash (b, i, \alpha, A\beta)$$

Dacă au fost epuizate toate alternativele pentru A și se renunță la expandarea acestui nod.

Apliiție.

Fie gramatica expresiilor aritmetice simple, fără recursie stângă:

1. $E \rightarrow TA$
2. $A \rightarrow +TA$
3. $A \rightarrow \varepsilon$
4. $T \rightarrow FB$
5. $B \rightarrow *FB$
6. $B \rightarrow \varepsilon$
7. $F \rightarrow a$
8. $F \rightarrow (E)$

Să se determine analiza sintactică la stânga pentru $w = a + a$.

$$\begin{aligned} & (q, 1, \varepsilon, E\$) \vdash (q, 1, E_1, TAS) \vdash (q, 1, E_1 T_1, FBAS) \vdash (q, 1, E_1 T_1 F_1, aBAS) \vdash (q, 2, E_1 T_1 F_1 a, BAS) \vdash \\ & \vdash (q, 2, E_1 T_1 F_1 a B_1, *FBAS) \vdash (q, 3, E_1 T_1 F_1 a B_1 *, FBAS) \vdash (q, 3, E_1 T_1 F_1 a B_1 * F_1, aBAS) \vdash \\ & \vdash (q, 4, E_1 T_1 F_1 a B_1 * F_1 a, BAS) \vdash (q, 4, E_1 T_1 F_1 a B_1 * F_1 a B_1, *FBAS) \vdash (b, 4, E_1 T_1 F_1 a B_1 * F_1 a B_1, *FBAS) \vdash \\ & \vdash (q, 4, E_1 T_1 F_1 a B_1 * F_1 a B_2, AS) \vdash (q, 4, E_1 T_1 F_1 a B_1 * F_1 a B_2 A_1, +TAS) \vdash (b, 4, E_1 T_1 F_1 a B_1 * F_1 a B_2 A_1, +TAS) \vdash \\ & \vdash (q, 4, E_1 T_1 F_1 a B_1 * F_1 a B_2 A_2, \$) \vdash (t, 4, E_1 T_1 F_1 a B_1 * F_1 a B_2 A_2, \varepsilon) \end{aligned}$$

$$h(E_1T_1F_1aB_1 * F_1aB_2A_2) = 1475763$$

Teme propuse:

1. Implementați algoritmul de analiză sintactică descendentă care utilizează STOS
2. Implementați algoritmul de analiză sintactică descendentă care utilizează un traducător finit.
3. Implementați algoritmul de analiză sintactică descendentă cu reveniri.