

Fordítóprogramok

3. előadás

Aszalós László

2015. szeptember 29.

- G grammatika
- L nyelv
- $L(G)$ a G grammatika által generált nyelv
- (N, T, P, S) grammatika (V_N, V_T, H, S) helyett
- \rightarrow helyettesítés
- $A \rightarrow \alpha|\beta$ két azonos oldalú helyettesítési szabály
- $\Rightarrow, \xRightarrow{*}, \xRightarrow{+}$ levezetés, $\geq 0, \geq 1$ alkalmazás
- a, b, c terminális szimbólumok (T)

- A, B, C nemterminális szimbólumok (N)
- X, Y, Z terminális vagy nemterminális szimbólumok ($T \cup N$)
- α, β, γ terminális vagy nemterminális szimbólumok sorozata $((T \cup N)^*)$
- x, y, z terminális szimbólumok sorozata (T^*)

Alapfogalmak

Definíció

Legyen $G = (N, T, P, S)$ egy grammatika.

- Ha $S \xRightarrow{*} \alpha$, akkor az α -t *mondatformának* nevezzük.
- Ha $S \xRightarrow{*} x$, akkor az x a grammatika által definiált nyelv egy mondata.

Definíció.

Legyen a $G = (N, T, P, S)$ grammatikának $\alpha = \alpha_1\beta\alpha_2$ egy mondatformája.

- A β -t az α egy *részmondatának* nevezzük, ha van olyan A szimbólum, amelyre $S \xRightarrow{*} \alpha_1 A \alpha_2$ és $A \xRightarrow{+} \beta$.
- A β egy *egyszerű részmondata* α -nak, ha a fentiekben az $A \Rightarrow \beta$ teljesül, azaz $A \rightarrow \beta \in P$.

$$E \rightarrow T \mid E + T, T \rightarrow F \mid T * F, F \rightarrow i \mid (E)$$

Mondatforma: $E + T * i + T * F$

	részmondat	egyszerű részmondat
$i + T$		
$i + T * F$		
$E + T * i$		
$T * i$		
$T * F$		

Mondatforma: $E + T * i + T * F$

	részmondat	egyszerű részmondat
$i + T$	-	-
$i + T * F$	-	-
$E + T * i$	+	-
$T * i$	+	-
$T * F$	+	+

- Az elemzés során a kapott mondat **szintaxisfáját** kívánjuk elkészíteni.
- Ennek levelei a grammatika terminális szimbólumai, a többi csúcs nemterminális szimbólumoknak felel meg, míg a gyökérelem a grammatika kezdőszimbóluma.
- A mondatforma szintaxisfájában a levelekhez nemterminális szimbólumok is tartozhatnak.

Feltételeink

- Csak olyan grammatikákat vizsgálunk, ahol minden mondathoz csak 1 szintaxisfa létezik (egyértelmű grammatikák).
- a grammatika ciklusmentes ($A \xRightarrow{+} A$ nem lehetséges)
- és redukált (nincs benne felesleges nemterminális):
 - ▶ minden szimbólum előfordul valamely levezetésben,
 - ▶ minden nemterminálisból levezethető valamilyen terminális sorozat)

Definíció

- Egy mondatforma legbaloldalibb egyszerű részmondatát a mondatforma *nyelének* nevezzük.
- Ha az $S \xRightarrow{*} x$ levezetésben minden helyettesítés a legbaloldalibb helyettesítés, akkor ezt a levezetést *legbaloldalibb levezetésnek* nevezzük.

Felülről-lefele elemzés

Definíció.

- Ha $A \rightarrow \alpha \in P$, akkor az $xA\beta$ mondatforma *legbaloldalibb helyettesítése* $x\alpha\beta$.
- $S \xRightarrow{*} x$ levezetésben minden helyettesítés a legbaloldalibb helyettesítés, akkor ezt a levezetést *legbaloldalibb levezetésnek* nevezzük

Példa

Készítse el az $i*(i+i)+i$ legbaloldalibb levezetését!

Tétel

Ha $S \xRightarrow{*} x\alpha \xRightarrow{*} yz$, ahol $|x| = |y|$, akkor $x = y$.

Definíció.

- A G grammatikában az A szimbólumot *balrekurzív szimbólumnak* nevezzük, ha $A \xRightarrow{+} A\alpha$.
- Az A *közvetlen balrekurzív szimbólum*, ha az $A \rightarrow A\alpha$ teljesül.
- Ha G tartalmaz legalább egy balrekurzív szimbólumot, akkor G -t balrekurzív grammatikának nevezzük.

Megjegyzés

A balrekurzivitás megszüntethető: az $A \rightarrow A\alpha|\beta$ szabályok helyettesíthetők az $A \rightarrow \beta|\beta A'$ és $A' \rightarrow \alpha|\alpha A'$ szabályokkal.

Teljes visszalépéses elemző algoritmus

Aho és Ullman, 1972

- Ha a grammatikában vannak $A \rightarrow \alpha_1 | \dots | \alpha_k$ szabályok, akkor az alternatívákat indexekkel különböztessük meg: $A_1 \rightarrow \alpha_1, \dots, A_k \rightarrow \alpha_k$!
- Jelölje c_1, \dots, c_n az input karaktereit, és jelölje $c_n + 1$ az inputzáró $\#$ jelet!
- Az elemzés állapotát (s, i, α, β) négyes jelöli, ahol
 - ▶ s az állapot típusa (q normál állapot, b visszalépés, t elemzés vége),
 - ▶ i az aktuális karakterre mutató pointer,
 - ▶ α a vizsgált mondatforma elemzésének történetét tároló verem,
 - ▶ β a mondatformát tartalmazó verem.
- A kezdőállapot $(q, 1, \lambda, S)$
- az állapotátmenetet $(s, i, \alpha, \beta) \mapsto (t, j, \gamma, \delta)$ jelzi.

Állapotátmentek

- ① Ha $A_1 \rightarrow \gamma_1$, akkor $(q, i, \alpha, A\beta) \mapsto (q, i, \alpha A_1, \gamma_1\beta)$
- ② Ha $c_i = a$, akkor $(q, i, \alpha, a\beta) \mapsto (q, i + 1, \alpha a, \beta)$
- ③ $(q, n + 1, \alpha, \lambda) \mapsto (t, n + 1, \alpha, \lambda)$
- ④ Ha $c_i \neq a$, akkor $(q, i, \alpha, a\beta) \mapsto (b, i, \alpha, a\beta)$
- ⑤ $(b, i, \alpha a, \beta) \mapsto (b, i - 1, \alpha, a\beta)$
- ⑥ ha az aktuális állapot $(b, i, \alpha A_j, \gamma_j\beta)$
 - ① $i = 1$, $A = S$, és S -nek nincs több helyettesítési szabálya. Ekkor nincs átmenet, hibával leáll az algoritmus
 - ② Van $A_{j+1} \rightarrow \gamma_{j+1}$ szabály: $(b, i, \alpha A_j, \gamma_j\beta) \mapsto (q, i, \alpha A_{j+1}, \gamma_{j+1}\beta)$
 - ③ egyébként ha felhasználtuk A minden alternatíváját:
 $(b, i, \alpha A_j, \gamma_j\beta) \mapsto (b, i, \alpha, A\beta)$

Megjegyzés

Ha az elemzés $(t, n + 1, \alpha, \lambda)$ állapottal fejeződik be, akkor α alapján megkonstruálható a legbaloldali levezetés.

Feladat

Elemezzük az $i + i$ inputot az

- $E \rightarrow T \mid TE'$
- $E' \rightarrow +T \mid +TE'$
- $T \rightarrow F \mid FT'$
- $T' \rightarrow *F \mid *FT'$ és
- $F \rightarrow i \mid (E)$

szabályokat tartalmazó grammatikában!

Megoldás

$(q, 1, \lambda, E) \mapsto (q, 1, E_1, T) \mapsto (q, 1, E_1 T_1, F) \mapsto (q, 1, E_1 T_1 F_1, i) \mapsto$
 $(q, 2, E_1 T_1 F_1 i, \lambda) \mapsto (b, 2, E_1 T_1 F_1 i, \lambda) \mapsto (b, 1, E_1 T_1 F_1, i) \mapsto$
 $(q, 1, E_1 T_1 F_2, (E)) \mapsto (b, 1, E_1 T_1 F_2, (E)) \mapsto (b, 1, E_1 T_1, F) \mapsto$
 $(q, 1, E_1 T_2, FT') \mapsto (q, 1, E_1 T_2 F_1, iT') \mapsto (q, 2, E_1 T_2 F_1 i, T') \mapsto$
 $(q, 2, E_1 T_2 F_1 iT'_1, *F) \mapsto (b, 2, E_1 T_2 F_1 iT'_1, *F) \mapsto$
 $(q, 2, E_1 T_2 F_1 iT'_2, *FT') \mapsto (b, 2, E_1 T_2 F_1 iT'_2, *FT') \mapsto$
 $(b, 2, E_1 T_2 F_1 i, T') \mapsto (b, 1, E_1 T_2 F_1, iT') \mapsto (q, 1, E_1 T_2 F_2, (E)T') \mapsto$
 $(b, 1, E_1 T_2 F_2, (E)T') \mapsto (b, 1, E_1 T_2, FT') \mapsto (b, 1, E_1, T) \mapsto$
 $(q, 1, E_2, TE') \mapsto (q, 1, E_2 T_1, FE') \mapsto (q, 1, E_2 T_1 F_1, iE') \mapsto$
 $(q, 2, E_2 T_1 F_1 i, E') \mapsto (q, 2, E_2 T_1 F_1 iE'_1, +T) \mapsto (q, 3, E_2 T_1 F_1 iE'_1 +, T) \mapsto$
 $(q, 3, E_2 T_1 F_1 iE'_1 + T_1, F) \mapsto (q, 3, E_2 T_1 F_1 iE'_1 + T_1 F_1, i) \mapsto$
 $(q, 4, E_2 T_1 F_1 iE'_1 + T_1 F_1 i, \lambda)$