

Μεταγλωττιστές

Γιώργος Δημητρίου

Μάθημα 4^ο

Συντακτική Ανάλυση

-
-
- Επαλήθευση της σύνταξης του προγράμματος
- Κατασκευή συντακτικού δέντρου
- Η κεντρική φάση της “Μετάφρασης Οδηγούμενης από τη Σύνταξη”
 - Από εδώ ξεκινά η παραγωγή ενδιάμεσου κώδικα
- Μηνύματα ελέγχου λαθών
- Εισαγωγή/Προσθήκη πληροφορίας στον ΠΣ

Μέθοδοι Συντακτικής Ανάλυσης

- Ντετερμινιστικές / Μη ντετερμινιστικές μέθοδοι
 - Διφορούμενες γραμματικές
- Από πάνω προς τα κάτω
 - ΣΑ πρόβλεψης της επόμενης λεκτικής μονάδας
- Από κάτω προς τα πάνω
 - ΣΑ ολίσθησης/ελάττωσης

Ορισμοί: Σύνολο FIRST

-
-
- $G = (T, N, P, S)$
- Έστω $\alpha \in (T \cup N)^*$
- Το σύνολο $\text{FIRST}(\alpha) \subseteq T \cup \{\varepsilon\}$ ορίζεται ως εξής:
- Εάν $\alpha \Rightarrow^* a\beta$, $a \in T$, $\beta \in (T \cup N)^*$ τότε $a \in \text{FIRST}(\alpha)$
- Εάν $\alpha \Rightarrow^* \varepsilon$, τότε $\varepsilon \in \text{FIRST}(\alpha)$

Ορισμοί: Σύνολο FOLLOW

- $G = (T, N, P, S)$
- Έστω $A \in N$
- Το σύνολο $FOLLOW(A) \subseteq T \cup \{EOF\}$ ορίζεται ως εξής:
- Εάν $S \Rightarrow^* \alpha A a \beta$, $\alpha, \beta \in (T \cup N)^*$, $a \in T$, τότε $a \in FOLLOW(A)$
- Εάν $S \Rightarrow^* \alpha A$, $\alpha \in (T \cup N)^*$ τότε $EOF \in FOLLOW(A)$

Υπολογισμός Συνόλων FIRST

- Αρχικοποίηση $FIRST(A)$, για κάθε $A \in N$
- Αρχικοποίηση $FIRST(\alpha)$, για κάθε επίθεμα $\alpha \in (T \cup N)^*$ δεξιών μελών των κανόνων της G
- Επανάλαβε
 - $\forall A \rightarrow \alpha : FIRST(A) := FIRST(A) \cup FIRST(\alpha)$
 - \forall επίθεμα $\alpha = A\beta$, όπου $\beta \in (T \cup N)^*$:
 - $FIRST(\alpha) := FIRST(\alpha) \cup (FIRST(A) - \{\epsilon\})$
 - $\text{Αν } \epsilon \in FIRST(A) : FIRST(\alpha) := FIRST(\alpha) \cup FIRST(\beta)$

μέχρι να μην υπάρχει μεταβολή στα σύνολα

Υπολογισμός Συνόλων FOLLOW

- Αρχικοποίηση $FOLLOW(A)$, για κάθε $A \in N$
 - Επανάλαβε
 - $\forall A \rightarrow \alpha B \beta$, όπου $A, B \in N$, $\alpha, \beta \in (T \cup N)^*$:
 - $FOLLOW(B) := FOLLOW(B) \cup (FIRST(\beta) - \{\epsilon\})$
 - Αν $\epsilon \in FIRST(\beta)$:
 $FOLLOW(B) := FOLLOW(B) \cup FOLLOW(A)$
- μέχρι να μην υπάρχει μεταβολή στα σύνολα

Παράδειγμα

- Για τη γραμματική:

$$E \rightarrow T E'$$

$$E' \rightarrow + T E' \mid \varepsilon$$

$$T \rightarrow F T'$$

$$T' \rightarrow * F T' \mid \varepsilon$$

$$F \rightarrow (E) \mid id$$

Παράδειγμα

- Σύνολα FIRST:

	Αρχικοποίηση	1η επανάληψη	2 ^η επανάληψη	3η επανάληψη	4η επανάληψη
E	\emptyset	\emptyset	\emptyset	{ (, id }	
E'	\emptyset	{ +, ϵ }			
T	\emptyset	\emptyset	{ (, id }		
T'	\emptyset	{ *, ϵ }			
F	\emptyset	{ (, id }			
T E'	\emptyset	\emptyset	{ (, id }		
+ T E'	{ + }				
F T'	\emptyset	{ (, id }			
* F T'	{ * }				
(E)	{ (}				
E)	\emptyset	\emptyset	\emptyset	{ (, id }	
)	{) }				
id	{ id }				

Παράδειγμα

- Σύνολα FOLLOW:

	Αρχικοποίηση	1 ^η επανάληψη	2 ^η επανάληψη
E	{ EOF }	{ EOF,) }	
E'	∅	{ EOF,) }	
T	∅	{ +, EOF,) }	
T'	∅	{ +, EOF,) }	
F	∅	{ *, +, EOF,) }	

ΣΑ από Πάνω προς τα Κάτω

-
-
- Ξεκίνημα με το αρχικό σύμβολο της γραμματικής
- Επίλεξε έναν κανόνα με αριστερό μέλος το τρέχον σύμβολο **ανάλογα με την είσοδο**
- Κατασκεύασε ένα δέντρο με ρίζα το σύμβολο αυτό και παιδιά τα σύμβολα του δεξιού μέλους του κανόνα
- Επανάλαβε για κάθε μη τερματικό από τα σύμβολα αυτά

ΣΑ LL(k)

- Σειρά επεξεργασίας των μη τερματικών συμβόλων από το δεξιό μέλος
 - Από αριστερά προς τα δεξιά
- Επιλογή κανόνα
 - Με βάση τις επόμενες k λεκτικές μονάδες
- Ανάγνωση εισόδου από αριστερά προς τα δεξιά

Γραμματικές LL(1)

-
-
- Για να είναι μια γραμματική LL(1), πρέπει να μην έχει συγκρούσεις FIRST/FIRST και FIRST/FOLLOW :
- Για δύο κανόνες $A \rightarrow \alpha$ και $A \rightarrow \beta$, $\alpha \neq \beta$:
$$\text{FIRST}(\alpha) \cap \text{FIRST}(\beta) = \emptyset$$
- Για ένα μη τερματικό σύμβολο $A \Rightarrow^* \varepsilon$:
$$\text{FIRST}(A) \cap \text{FOLLOW}(A) = \emptyset$$

Γραμματικές LL(1)

- Συγκρούσεις FIRST/FIRST και FIRST/FOLLOW προκύπτουν **συνήθως** από:
- Αριστερή αναδρομή (άμεση ή έμμεση)
- Δύο εναλλακτικούς κανόνες με κοινό πρόθεμα στα δεξιά μέλη τους
- Δύο εναλλακτικούς κανόνες τα δεξιά μέλη των οποίων παράγουν την κενή συμβολοσειρά

Μετατροπή Γραμματικής σε LL(1)

-
-
- Μια γραμματική που δεν είναι LL(1) **πιθανά** να μπορεί να μετατραπεί σε LL(1) με μια σειρά μετασχηματισμών:
- Αριστερή παραγοντοποίηση
- Απαλοιφή αριστερής αναδρομής
- Αντικατάσταση
- Η μετατροπή δεν είναι εφικτή για όλες τις γραμματικές που δεν είναι LL(1)

Αριστερή Παραγοντοποίηση

- Εάν υπάρχουν κανόνες της μορφής:
$$A \rightarrow \alpha\beta_1 \mid \dots \mid \alpha\beta_n$$
- τότε οι κανόνες μετασχηματίζονται ως εξής:
$$A \rightarrow \alpha B$$
$$B \rightarrow \beta_1 \mid \dots \mid \beta_n$$
- Μη προφανή κοινά προθέματα φανερώνονται με αντικατάσταση

Απαλοιφή Αριστερής Αναδρομής

-
-
- Αν υπάρχουν κανόνες της μορφής:
$$A \rightarrow A\alpha_1 \mid \dots \mid A\alpha_n \mid \beta_1 \mid \dots \mid \beta_m$$
- τότε η αναδρομή μετασχηματίζεται ως εξής:
$$A \rightarrow \beta_1 B \mid \dots \mid \beta_m B$$
$$B \rightarrow \alpha_1 B \mid \dots \mid \alpha_n B \mid \varepsilon$$
- Έμμεση αναδρομή μετασχηματίζεται εύκολα σε άμεση με αντικατάσταση

ΣΑ Αναδρομικής Κατάβασης

-
-
- Μια ρουτίνα για την ανάλυση κάθε μη τερματικού συμβόλου A με $A \rightarrow \alpha_1 \mid \dots \mid \alpha_n$:
- $\forall i = 1, \dots, n$: αν $\text{token} \in \text{FIRST}(\alpha_i)$ τότε
 - κώδικας για την αναγνώριση της α_i
- αλλιώς αν $\varepsilon \notin \text{FIRST}(\alpha_1) \cup \dots \cup \text{FIRST}(\alpha_n)$
 - συντακτικό σφάλμα
- αλλιώς αν $\text{token} \notin \text{FOLLOW}(A)$
 - συντακτικό σφάλμα

ΣΑ Αναδρομικής Κατάβασης

- Ο κώδικας αναγνώρισης μιας συμβολοσειράς (από αριστερά προς τα δεξιά) περιλαμβάνει:
 - Κλήση ρουτίνας για μη τερματικό σύμβολο
 - Έλεγχο ταύτισης με την επόμενη λεκτική μονάδα (ΛΜ) στην είσοδο για τερματικό σύμβολο
 - Τίποτα (έλεγχο με το FOLLOW του αριστερού μέλους) για κενή συμβολοσειρά
- Παράθεση κώδικα για παράθεση συμβόλων
- Εύκολος χειρισμός τελεστή * σε κανόνες

Παράδειγμα

- Για τη γραμματική:

$$E \rightarrow T (+ T)^*$$

$$T \rightarrow F (* F)^*$$

$$F \rightarrow (E) \mid \text{id}$$

καλούμε το ΣΑ ως εξής:

token := επόμενη ΛΜ
κάλεσε τη ρουτίνα για το E
αν token ≠ EOF τότε
συντακτικό σφάλμα

Παράδειγμα

- Για τον κανόνα:

$$E \rightarrow T (+ T)^*$$

η ρουτίνα ΣΑ θα είναι:

κάλεσε τη ρουτίνα για το T
όσο token = + κάνε
token := επόμενη ΛΜ
κάλεσε τη ρουτίνα για το T
τέλος βρόχου

Παράδειγμα

- Για τον κανόνα:

$$T \rightarrow F (* F)^*$$

η ρουτίνα ΣΑ θα είναι:

κάλεσε τη ρουτίνα για το F
όσο token = * κάνε
token := επόμενη ΛΜ
κάλεσε τη ρουτίνα για το F
τέλος βρόχου

Παράδειγμα

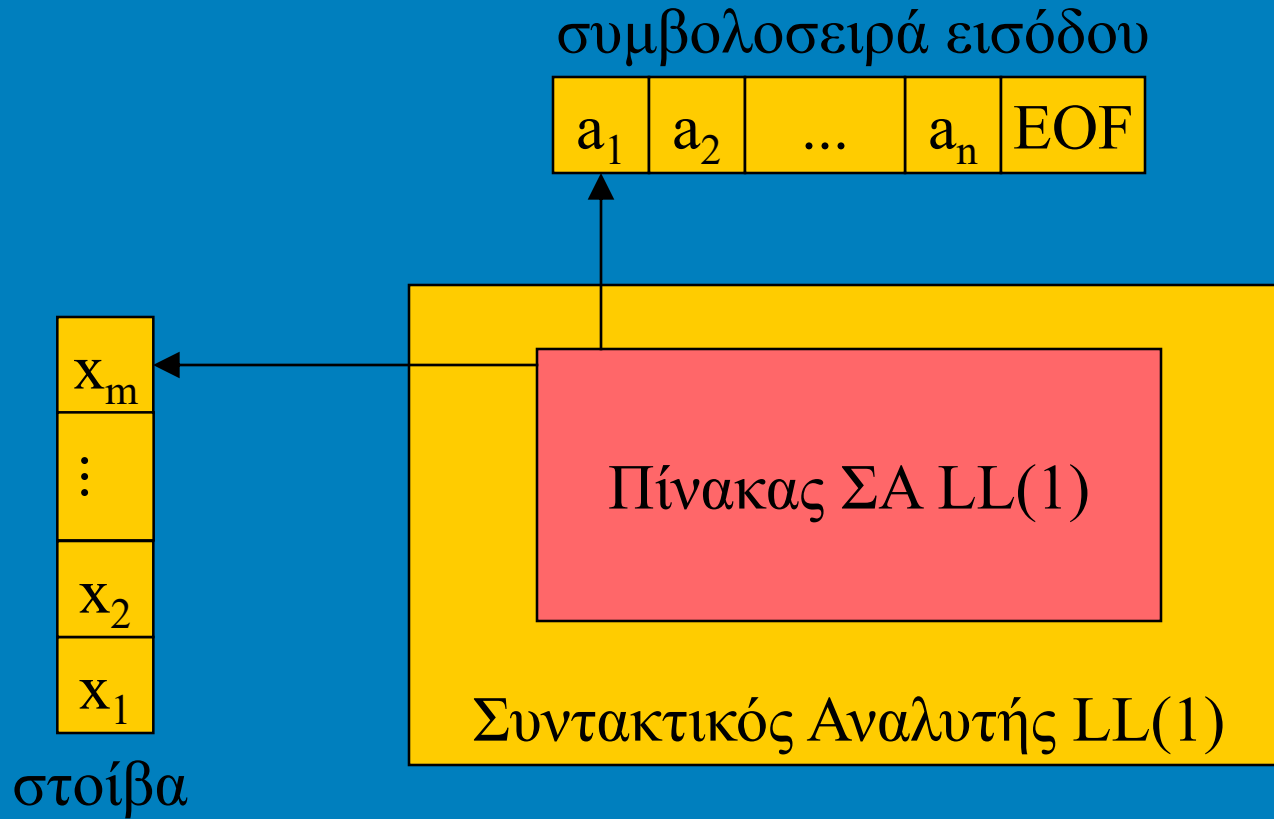
- Για τους κανόνες:

$$F \rightarrow (E) \mid id$$

η ρουτίνα ΣΑ θα είναι:

*αν token = (τότε
token := επόμενη ΛΜ
κάλεσε τη ρουτίνα για το E
αν token =) τότε
token := επόμενη ΛΜ
αλλιώς συντακτικό σφάλμα
αλλιώς αν token = id τότε
token := επόμενη ΛΜ
αλλιώς συντακτικό σφάλμα*

ΣΑ LL(1) με χρήση ΑΣ



Πίνακας ΣΑ LL(1)

- Αρχικοποίηση M
- Για κάθε κανόνα $A \rightarrow \alpha$
 - $\forall a \in \text{FIRST}(\alpha) : M(A,a) := A \rightarrow \alpha$
 - Εάν $\epsilon \in \text{FIRST}(\alpha)$ τότε:
 - $\forall a \in \text{FOLLOW}(A) : M(A,a) := A \rightarrow \alpha$
 - Εάν $\text{EOF} \in \text{FOLLOW}(A) : M(A,\text{EOF}) := A \rightarrow \alpha$

Λειτουργία ΑΣ LL(1)

- $H = N \cup T, h_0 = S$
- Κινήσεις ΑΣ:
 - Αν η κορυφή της στοίβας περιέχει το ίδιο σύμβολο με το επόμενο σύμβολο εισόδου, τότε αυτό αφαιρείται και προχωράμε κατά ένα σύμβολο
 - Αν η κορυφή της στοίβας περιέχει το μη τερματικό σύμβολο A , το επόμενο σύμβολο εισόδου είναι το a , και ορίζεται το στοιχείο $M(A,a)$, τότε το A αντικαθίσταται με το δεξί μέλος του κανόνα $M(A,a)$, το οποίο εισάγεται στη στοίβα αντίστροφα
 - Σε κάθε άλλη περίπτωση έχουμε σφάλμα

Παράδειγμα

- Για τη γραμματική:

$$E \rightarrow T E'$$

$$E' \rightarrow + T E' \mid \varepsilon$$

$$T \rightarrow F T'$$

$$T' \rightarrow * F T' \mid \varepsilon$$

$$F \rightarrow (E) \mid \text{id}$$

Παράδειγμα

- Πίνακας ΣΑ LL(1):

	id	+	*	()	EOF
E	$E \rightarrow TE'$			$E \rightarrow TE'$		
E'		$E' \rightarrow +TE'$			$E' \rightarrow \epsilon$	$E' \rightarrow \epsilon$
T	$T \rightarrow FT'$			$T \rightarrow FT'$		
T'		$T' \rightarrow \epsilon$	$T' \rightarrow *FT'$		$T' \rightarrow \epsilon$	$T' \rightarrow \epsilon$
F	$F \rightarrow id$			$F \rightarrow (E)$		

Παράδειγμα

- Αναγνώριση συμβολοσειράς “id + id * id”:

βήμα	στοίβα	είσοδος	κίνηση
0	E	id + id * id \$	κανόνας $E \rightarrow T E'$
1	E' T	id + id * id \$	κανόνας $T \rightarrow F T'$
2	E' T' F	id + id * id \$	κανόνας $F \rightarrow id$
3	E' T' id	id + id * id \$	απορρόφηση id
4	E' T'	+ id * id \$	κανόνας $T' \rightarrow \epsilon$
5	E'	+ id * id \$	κανόνας $E' \rightarrow + T E'$
6	E' T +	+ id * id \$	απορρόφηση +
7	E' T	id * id \$	κανόνας $T \rightarrow F T'$
8	E' T' F	id * id \$	κανόνας $F \rightarrow id$
9	E' T' id	id * id \$	απορρόφηση id
10	E' T'	* id \$	κανόνας $T' \rightarrow * F T'$
11	E' T' F *	* id \$	απορρόφηση *
12	E' T' F	id \$	κανόνας $F \rightarrow id$
13	E' T' id	id \$	απορρόφηση id
14	E' T'	\$	κανόνας $T' \rightarrow \epsilon$
15	E'	\$	κανόνας $E' \rightarrow \epsilon$
16		\$	αποδοχή