



Σχεδίαση Γλωσσών & Μεταγλωττιστές

Ενότητα 5: Καθοδική Συντακτική Ανάλυση –
Αλγόριθμος LL(1)

Επ. Καθ. Π. Κατσαρός
Τμήμα Πληροφορικής



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδεια χρήσης, η άδεια χρήσης αναφέρεται ρητώς.



Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «Ανοικτά Ακαδημαϊκά Μαθήματα στο Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Καθοδική Ανάλυση

- Ξεκινά από τη αρχή της γραμματικής (ρίζα δένδρου) για να αναπτύξει τη συμβολοσειρά, που αναλύει, αν αυτή αποτελεί πρόταση.
- Καμία από τις τεχνικές που θα εξετάσουμε δεν μπορεί να εφαρμοσθεί σε αριστερά αναδρομικές γραμματικές.
- Διακρίνουμε της τεχνικές της οπισθοδρόμησης και τις τεχνικές της πρόγνωσης.



Καθοδική ανάλυση με οπισθοδρόμηση I

1. Δοθέντος συγκεκριμένου μη τερματικού συμβόλου εφαρμόζεται ο πρώτος κανόνας της γραμματικής.
2. Στην προτασιακή μορφή, που προκύπτει επιλέγεται το πρώτο από αριστερά μη τερματικό σύμβολο και εφαρμόζεται ο πρώτος κανόνας, που αναφέρεται σε αυτό.
3. Γίνεται επαναληπτική εφαρμογή του βήματος 2, για κάθε ένα από τα μη τερματικά σύμβολα, που ακολουθούν, μέχρι τη στιγμή της παραγωγής,
 - μιας σειράς τερματικών συμβόλων (πρόταση γλώσσας αν ταυτίζεται) ή
 - τμήματος τερματικών συμβόλων της προτασιακής μορφής, που διαφέρει από το αντίστοιχο τμήμα της συμβολοσειράς εισόδουΗ δεύτερη περίπτωση μπορεί να είναι αποτέλεσμα εσφαλμένης επιλογής κανόνα. Αναίρειται ο κανόνας που εφαρμόστηκε τελευταίος και χρησιμοποιείται ο επόμενος που ισχύει για το ίδιο μη τερματικό σύμβολο. Αν έχουν χρησιμοποιηθεί όλοι οι κανόνες, τότε ο αλγόριθμος διατηρεί το σύμβολο ως έχει και προχωρά στην αναίρεση του κανόνα για το προηγούμενο μη τερματικό σύμβολο και οπισθοδρομεί στην κατάλληλη θέση της συμβολοσειράς εισόδου.



Καθοδική ανάλυση με οπισθοδρόμηση II

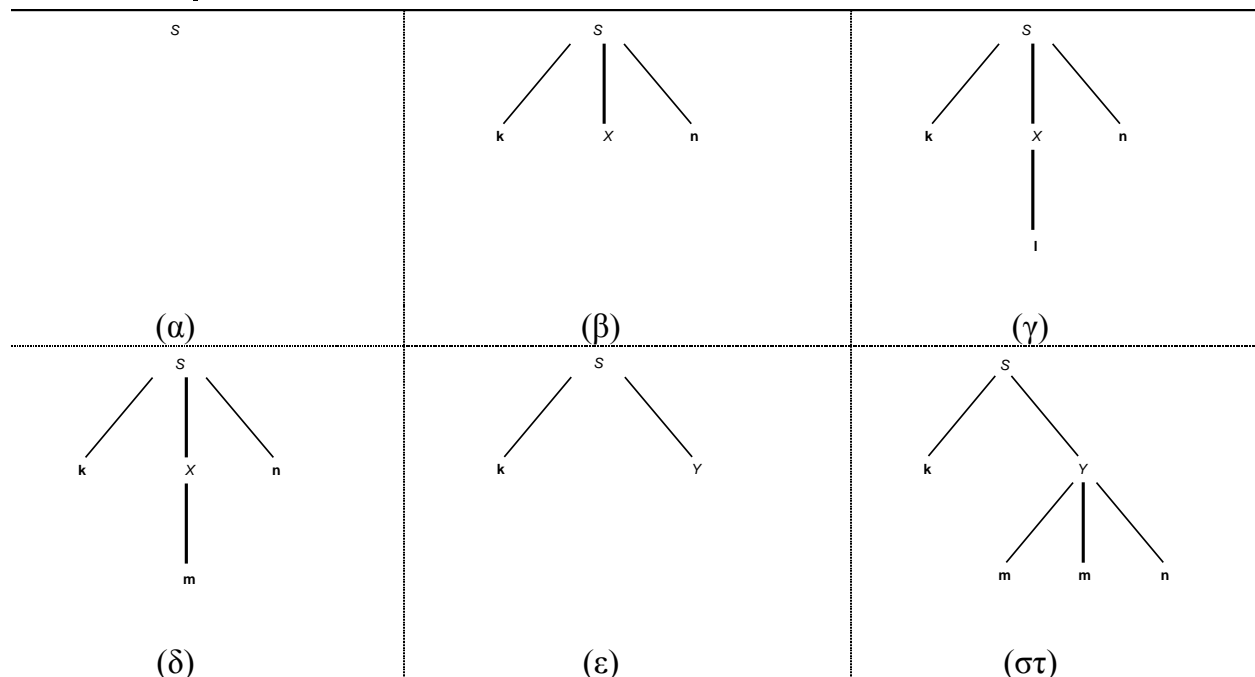
ΠΑΡΑΔΕΙΓΜΑ

$S = "k" X "n" \mid "k" Y.$

$X = "l" \mid "m".$

$Y = "mmn" \mid "nmm".$

ΠΡΟΤΑΣΗ: "kmmm"



Καθοδική ανάλυση με οπισθοδρόμηση III

- Ιδιαίτερα πολύπλοκη
- Όχι αποδοτική
- Αρκετά γενική, δηλ. αναγνωρίζει πιο πολλές γλώσσες από τις άλλες τεχνικές καθοδικής ανάλυσης
- Αργεί στον εντοπισμό των λαθών και έτσι η ανάνηψη γίνεται πολύ δύσκολη



Σύνολα FIRST για Συντακτική Ανάλυση I

ΣΥΝΟΛΑ FIRST

Για μία σειρά συμβόλων α της γραμματικής ορίζουμε **FIRST**(α) ως το σύνολο των αναγνωριστικών που εμφανίζονται στην αρχή των συμβολοσειρών που παράγονται από την α

Διακρίνουμε τις εξής περιπτώσεις για την προτασιακή μορφή α :

1. εάν α είναι ένα γραμματικό σύμβολο ή ϵ , τότε

εάν α είναι ένα τερματικό a , τότε $\text{First}(\alpha) = \text{First}(a) = \{a\}$

εάν α είναι το ϵ , τότε $\text{First}(\alpha) = \text{First}(\epsilon) = \{\epsilon\}$

εάν α είναι ένα μη-τερματικό X και $X \rightarrow \beta_1 \mid \beta_2 \mid \dots \mid \beta_k$, τότε $\text{First}(\alpha) = \text{First}(X) = \text{First}(\beta_1) \cup \text{First}(\beta_2) \cup \dots \cup \text{First}(\beta_k)$

2. εάν α είναι μια συμβολοσειρά γραμματικών συμβόλων $\alpha = X_1X_2\dots X_N$, τότε

$\text{First}(\alpha) = \{\}; j=0;$

REPEAT

$j=j+1;$

Προσθέτουμε στο $\text{First}(\alpha)$ το $\text{First}(X_j) - \{\epsilon\}$

UNTIL X_j δεν είναι απαλείψιμο (δεν παράγει το ϵ) ή $j=N$

Εάν $X_1X_2\dots X_N$, είναι απαλείψιμη, τότε προσθέτουμε το ϵ στο $\text{First}(\alpha)$



Σύνολα FIRST για Συντακτική Ανάλυση II

ΠΑΡΑΔΕΙΓΜΑ ΥΠΟΛΟΓΙΣΜΟΥ ΣΥΝΟΛΩΝ FIRST

S	=	έκφραση.
έκφραση	=	όρος υπ_όροι.
υπ_όροι	=	"+" όρος υπ_όροι
		"-" όρος υπ_όροι
		ε.
όρος	=	παράγοντας υπ_παραγ.
υπ_παραγ	=	"*" παράγοντας υπ_παραγ
		"/" παράγοντας υπ_παραγ
		ε.
παράγοντας	=	"("έκφραση ")"
		"αριθμός".

Σύμβολα	FIRST	Σύμβολα	FIRST	Συμβ/ρές	FIRST
S	{ "(", "αριθμός" }	παράγοντας	{ "(", "αριθμός" }	"+" όρος υπ_όροι	{ "+" }
έκφραση	{ "(", "αριθμός" }	αριθμός	{ "αριθμός" }	"-" όρος υπ_όροι	{ "-" }
υπ_όροι	{ ε, "+", "-" }	"+"	{ "+" }	"*" παράγοντας υπ_παραγ	
{ "*" }					
όρος	{ "(", "αριθμός" }	"-"	{ "-" }	"/" παράγοντας υπ_παραγ	{ "/" }
υπ_παραγ	{ ε, "*", "/" }	"*"	{ "*" }		
		"/"	{ "/" }		
		"("	{ "(" }		



Σύνολα FOLLOW για Συντακτική Ανάλυση I

ΣΥΝΟΛΑ FOLLOW ΓΙΑ ΜΗ ΤΕΡΜΑΤΙΚΑ ΣΥΜΒΟΛΑ ΓΡΑΜΜΑΤΙΚΗΣ

Για ένα μη τερματικό σύμβολο A ορίζουμε το **FOLLOW**(A) ως το σύνολο των τερματικών που μπορούν να εμφανισθούν δεξιά του A (ακολουθούν) σε κάποια προτασιακή μορφή

Έστω $G=(S, N, T, P)$, $A \in N$ και $\$$ το τέλος μιας συμβολοσειράς

1. εάν $A = S$, τότε $\$ \in \text{Follow}(A)$

2. $\forall p \in P$, και A περιλαμβάνεται στο δεξί μέρος του p με τη μορφή:

$B \rightarrow \gamma A \delta$

i. εάν δ αρχίζει από τερματικό σύμβολο c , τότε $c \in \text{Follow}(A)$

ii. εάν δ αρχίζει από μη-τερματικό σύμβολο, τότε προσθέτουμε το $\text{First}(\delta) - \{\epsilon\}$

iii. εάν $\delta = \epsilon$ ή $\delta \Rightarrow^* \epsilon$, τότε προσθέτουμε το $\text{Follow}(B)$



Σύνολα FOLLOW για Συντακτική Ανάλυση II

Παραδείγματα FOLLOW (γραμματική αριθμητικών εκφράσεων)

$$FOLLOW(\acute{\epsilon}\kappa\phi\rho\alpha\sigma\eta) = \{\$, "\}''\}$$

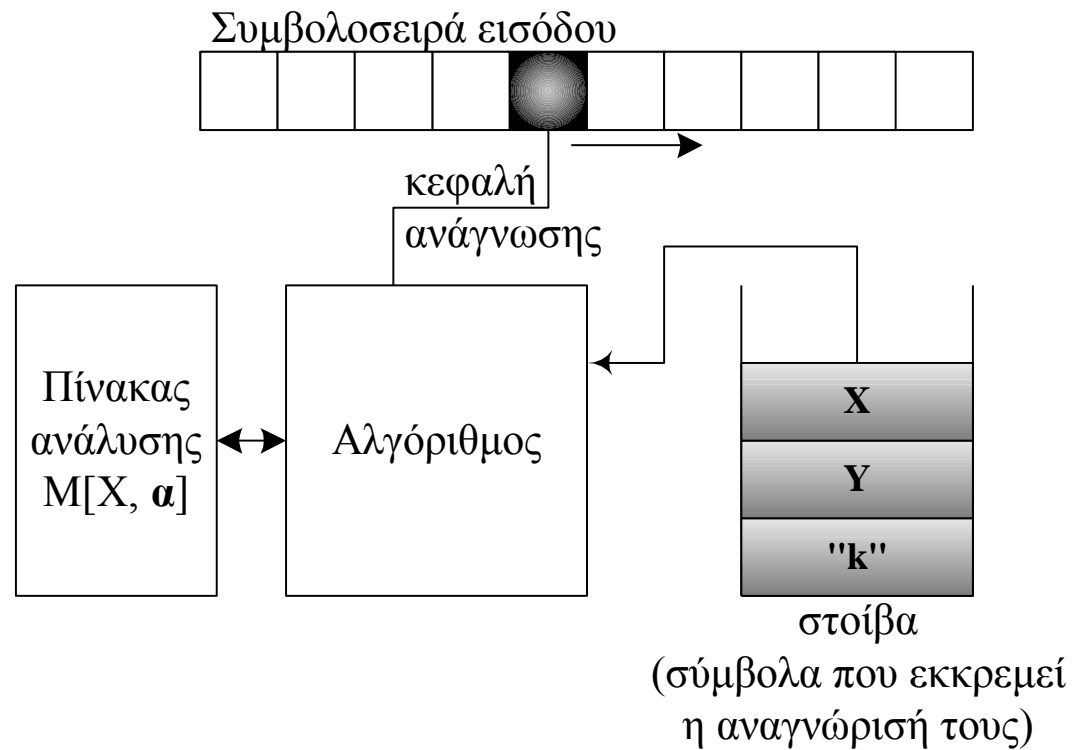
$$FOLLOW(\upsilon\pi_ό\rho\iota) = FOLLOW(\acute{\epsilon}\kappa\phi\rho\alpha\sigma\eta)$$

$$FOLLOW(ό\rho\varsigma) = FIRST(\upsilon\pi_ό\rho\iota) - \{\epsilon\} \cup FOLLOW(\acute{\epsilon}\kappa\phi\rho\alpha\sigma\eta) = \{ "+", "- ", "\$, "\}''\}$$

$$FOLLOW(\upsilon\pi_π\rho\alpha\rho\gamma) = FOLLOW(ό\rho\varsigma) = \{ "+", "- ", "\$, "\}''\}$$



Ανάλυση LL(1) I



Ανάλυση LL(1) III

ΠΑΡΑΔΕΙΓΜΑ (γραμματική αριθμητικών εκφράσεων)

S	=	έκφραση.
έκφραση	=	όρος υπ_όροι.
υπ_όροι	=	"+" όρος υπ_όροι "-" όρος υπ_όροι
		ε.
όρος	=	παράγοντας υπ_παραγ.
υπ_παραγ	=	"*" παράγοντας υπ_παραγ "/" παράγοντας υπ_παραγ
		ε.
παράγοντας	=	"("έκφραση ")" "αριθμός".

M	"αριθμός"	"+"	"-"	"*"	"/"	"(")"	\$
έκφραση	όρος υπ_όροι					όρος υπ_όροι		
υπ_όροι		"+" όρος υπ_όροι	"-" όρος υπ_όροι				ε	ε
όρος	παράγοντας υπ_παραγ					παράγοντας υπ_παραγ		
υπ_παραγ		ε	ε	"*" παράγοντας υπ_παραγ	"/" παράγοντας υπ_παραγ		ε	ε
παράγοντας	"αριθμός"					"(" έκφραση ")"		



Ανάλυση LL(1) IV

ΠΑΡΑΔΕΙΓΜΑ (γραμματική αριθμητικών εκφράσεων - συνέχεια)

LL(1) ανάλυση της συμβολοσειράς εισόδου $27-5*8$		
ΣΤΟΙΒΑ	ΣΥΜΒΟΛ.	ΠΑΡΑΓΩΓΗ
\$ έκφραση	27-5*8\$	
\$ υπ_όροι όρος	27-5*8\$	έκφραση = όρος υπ_όροι.
\$ υπ_όροι υπ_παραγ. παράγοντας	27-5*8\$	όρος = παράγοντας υπ_παραγ.
\$ υπ_όροι υπ_παραγ. “αριθμός”	27-5*8\$	παράγοντας = “αριθμός”.
\$ υπ_όροι υπ_παραγ.	-5*8\$	
\$ υπ_όροι	-5*8\$	υπ_παραγ. = ε.
\$ υπ_όροι όρος “-”	-5*8\$	υπ_όροι = “-” όρος υπ_όροι.
\$ υπ_όροι όρος	5*8\$	
\$ υπ_όροι υπ_παραγ. παράγοντας	5*8\$	όρος = παράγοντας υπ_παραγ.
\$ υπ_όροι υπ_παραγ. “αριθμός”	5*8\$	παράγοντας = “αριθμός”.
\$ υπ_όροι υπ_παραγ.	*8\$	
\$ υπ_όροι υπ_παραγ. παράγοντας “*”	*8\$	υπ_παραγ. = “*” παράγοντας υπ_παραγ.
\$ υπ_όροι υπ_παραγ. παράγοντας	8\$	
\$ υπ_όροι υπ_παραγ. “αριθμός”	8\$	παράγοντας = “αριθμός”.
\$ υπ_όροι υπ_παραγ.	\$	
\$ υπ_όροι	\$	υπ_παραγ. = ε.
\$	\$	υπ_όροι. = ε.



Ανάλυση LL(1) II

Αλγόριθμος υπολογισμού πίνακα ανάλυσης M

Για όλους τους κανόνες $A \rightarrow \alpha$, εκτελούνται τα ακόλουθα βήματα:

1. για κάθε τερματικό σύμβολο b , όπου $b \in \text{First}(\alpha)$, προσθέτουμε τον κανόνα $A \rightarrow \alpha$ στη θέση $M[A, b]$
2. εάν $\epsilon \in \text{First}(\alpha)$, τότε προσθέτουμε τον κανόνα $A \rightarrow \alpha$ στη θέση $M[A, c]$ για κάθε τερματικό σύμβολο c που περιέχεται στο σύνολο $\text{Follow}(A)$, επίσης προσθέτουμε τον κανόνα $A \rightarrow \alpha$ στη θέση $M[A, \$]$, εάν $\$ \in \text{Follow}(A)$

Τα στοιχεία του ΠΑ που δεν έχουν οριστεί με τα βήματα 1 και 2 θα έχουν τιμή `error`



Ανάλυση LL(1) V

Αλγόριθμος ανάλυσης LL(1)

```

push($);           // $ είναι το σύμβολο τέλους της συμβολοσειράς
push(S);           // S είναι η αρχή της γραμματικής
lookahead = get_next_token(); // διάβασε το επόμενο αναγνωριστικό
repeat
  X = top_of_stack();
  if (X είναι τερματικό ή X == $) then
    if (X == lookahead) then
      pop(X);
      lookahead = get_next_token();
    else error();
  else // X ένα μη τερματικό
    if ( M[X, lookahead] == X → Y1 Y2 ... Yk ) then
      pop(X);
      push(Yk); push(Yk-1); ... push(Y1);
    else error();
until (X = $)

```





Τέλος ενότητας

Επεξεργασία: Εμμανουέλα Στάχτιαρη
Θεσσαλονίκη, 21/07/2014



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΣΠΑ
2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ