

Μεταγλωττιστές

Γιώργος Δημητρίου

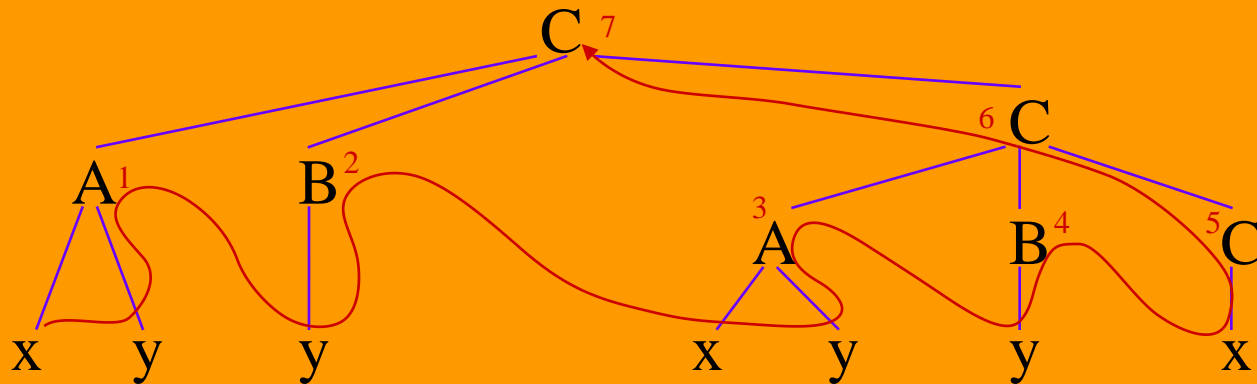
Μάθημα 5^ο

ΣΑ από Κάτω προς τα Πάνω

-
-
- Ξεκίνημα με την πρώτη λεκτική μονάδα
- Διάβασε διαδοχικές λεκτικές μονάδες αντικαθιστώντας το δεξί μέλος κάποιου κανόνα που έχει σχηματιστεί με το αριστερό του
- Κατασκεύασε ένα δέντρο με φύλλα τα σύμβολα του δεξιού μέλους του κανόνα και ρίζα το σύμβολο του αριστερού μέλους
- Επανέλαβε μέχρι να βρεθεί το αρχικό σύμβολο της γραμματικής

Παράδειγμα

- Για τη γραμματική:
 - $C \rightarrow ABC \mid x$
 - $A \rightarrow xy$
 - $B \rightarrow y$
- Μια ανάλυση από κάτω προς τα επάνω για αναγνώριση της εισόδου “xyyxyyx” είναι:



ΣΑ LR(k)

- Σειρά σχηματισμού των μη τερματικών συμβόλων στο δεξιό μέλος
 - Από αριστερά προς τα δεξιά (αντίστροφη της δεξιότερης παραγωγής)
- Επιλογή κανόνα και κίνησης
 - Με βάση τις επόμενες k λεκτικές μονάδες
- Ανάγνωση εισόδου από αριστερά προς τα δεξιά

Κινήσεις ΣΑ LR(k)

-
-
- Οι κινήσεις γίνονται με τη βοήθεια στοίβας
- Ολίσθηση
 - Ανάγνωση επόμενης λεκτικής μονάδας από την είσοδο και εισαγωγή αυτής στη στοίβα
- Ελάττωση
 - Αντικατάσταση στη στοίβα των συμβόλων του πλήρους δεξιού μέλους κάποιου κανόνα με το μη τερματικό σύμβολο στο αριστερό μέλος

Γραμματικές LR(0)

-
-
- Για να είναι μια γραμματική LR(0) πρέπει να μην έχει συγκρούσεις ολίσθησης/ελάττωσης και ελάττωσης/ελάττωσης:
 - Να μην υπάρχει επιλογή μεταξύ ολίσθησης και ελάττωσης για την ίδια κίνηση
 - Να μην υπάρχουν δύο ή περισσότεροι κανόνες που να μπορούν να χρησιμοποιηθούν στην ίδια κίνηση ελάττωσης

ΣΑ LR(0)

- Κατάσταση ΣΑ LR(0): Το σύνολο των προθεμάτων δεξιών μελών κανόνων που έχουν αναγνωριστεί σε κάποιο βήμα της ΣΑ
- Με την επόμενη κίνηση πρέπει να σχηματίζεται τουλάχιστον ένα νέο αποδεκτό πρόθεμα (viable prefix) δεξιού μέλους κανόνα
- Η ολίσθηση καταναλώνει μια λεκτική μονάδα
- Η ελάττωση είναι ε-μετάβαση και άρα δεν καταναλώνει λεκτική μονάδα

Η Έννοια του Στοιχείου

-
-
- Αναπαράσταση του προθέματος του δεξιού μέλους κάποιου κανόνα που έχει μέχρι τώρα αναγνωριστεί:
 - Ο κανόνας $A \rightarrow xAB$ δίνει τα στοιχεία
 - $A \rightarrow \bullet xAB$
 - $A \rightarrow x \bullet AB$
 - $A \rightarrow xA \bullet B$
 - $A \rightarrow xAB \bullet$

Άλλες Βοηθητικές Έννοιες

-
-
- Αν I είναι σύνολο στοιχείων μιας γραμματικής
- Συνάρτηση CLOSURE:
 1. $I \subseteq \text{CLOSURE}(I)$
 2. Αν $A \rightarrow \alpha \bullet B \beta \in \text{CLOSURE}(I)$, τότε για κάθε κανόνα $B \rightarrow \gamma : B \rightarrow \bullet \gamma \in \text{CLOSURE}(I)$
- Συνάρτηση GOTO:

Για κάθε σύμβολο $x \in T \cup N$ της γραμματικής

$$\text{GOTO}(I, x) = \text{CLOSURE}(\{A \rightarrow \alpha x \bullet \beta : A \rightarrow \alpha \bullet x \beta \in I\})$$

Κατασκευή Καταστάσεων LR(0)

-
-
- Διεύρυνση γραμματικής (αν χρειάζεται):
 $S' \rightarrow S$
- Αρχική κατάσταση $I_0 = \text{CLOSURE}(\{S' \rightarrow \bullet S\})$
- Οι καταστάσεις του ΣΑ LR(0) βρίσκονται από τις τιμές της συνάρτησης GOTO για κάθε νέα κατάσταση και κάθε σύμβολο της γραμματικής

Παράδειγμα

- Διευρύνουμε την προηγούμενη γραμματική:

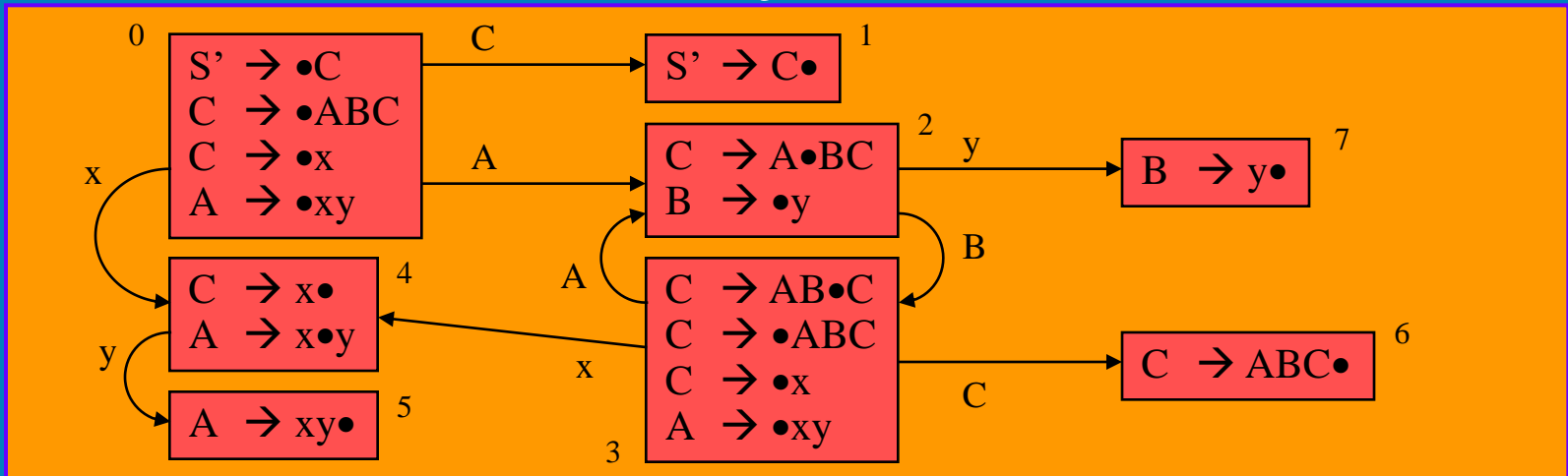
$$S' \rightarrow C^0$$

$$C \rightarrow ABC^1 \mid x^2$$

$$A \rightarrow xy^3$$

$$B \rightarrow y^4$$

οπότε οι καταστάσεις LR(0) θα είναι:



Πίνακες ΣΑ LR(0)

- Πίνακας ενέργειας ACTION
 - Ολίσθηση / Ελάττωση / Αποδοχή
 - Αν $A \rightarrow \alpha \bullet \alpha \beta \in I_i$ τότε $\text{ACTION}(i) = \text{ολίσθηση}$
 - Αν $A \rightarrow \alpha \bullet \in I_i$ και $A \neq S'$ τότε $\text{ACTION}(i) = \text{ελάττωση με τον κανόνα } A \rightarrow \alpha$
 - Αν $S' \rightarrow \alpha \bullet \in I_i$ τότε $\text{ACTION}(i) = \text{αποδοχή}$
- Πίνακας επόμενης κατάστασης NEXT
 - Για κάθε σύμβολο x της γραμματικής:
Αν $\text{GOTO}(I_i, x) = I_j$ τότε $\text{NEXT}(i, x) = j$

Παράδειγμα

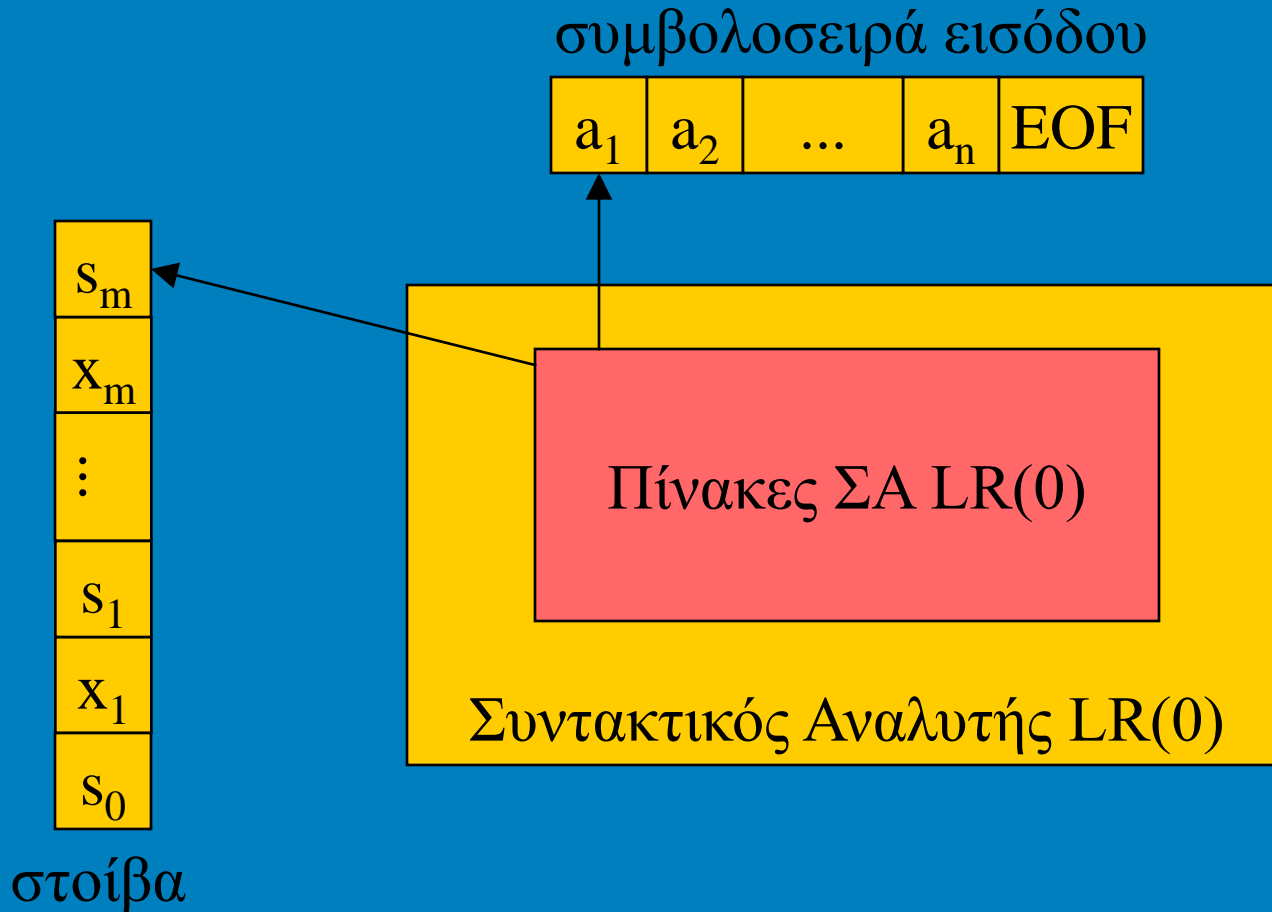
- Για την προηγούμενη διευρυμένη γραμματική:

| ACTION | |
|--------|-----------------------|
| 0 | ολίσθηση |
| 1 | αποδοχή |
| 2 | ολίσθηση |
| 3 | ολίσθηση |
| 4 | ολίσθηση / ελάττωση 2 |
| 5 | ελάττωση 3 |
| 6 | ελάττωση 1 |
| 7 | ελάττωση 4 |

| NEXT | | | | | |
|------|---|---|---|---|---|
| | x | y | A | B | C |
| 0 | 4 | | 2 | | 1 |
| 1 | | | | | |
| 2 | | 7 | | 3 | |
| 3 | 4 | | 2 | | 6 |
| 4 | | 5 | | | |
| 5 | | | | | |
| 6 | | | | | |
| 7 | | | | | |

- Η γραμματική δεν είναι LR(0)!

Λειτουργία ΣΑ LR(0)



Αλγόριθμος ΣΑ LR(0)

-
-
- Βάλε στη στοίβα την αρχική κατάσταση
- Επανέλαβε μέχρι αποδοχή ή απόρριψη:
 - Εκτέλεσε την ενέργεια για την κατάσταση στην κορυφή της στοίβας (τρέχουσα κατάσταση):
 - Ολίσθηση: βάλε στη στοίβα την επόμενη λεκτική μονάδα
 - Ελάττωση: βγάλε από τη στοίβα $2 \times$ τον αριθμό συμβόλων του δεξιού μέλους, θέσε ως τρέχουσα κατάσταση το σύμβολο στην κορυφή της στοίβας και βάλε σε αυτήν το αριστερό μέλος του κανόνα ελάττωσης
 - Βάλε στη στοίβα την επόμενη της τρέχουσας κατάστασης για το σύμβολο στην κορυφή της

Γραμματικές LR(k)

-
-
- Για να είναι μια γραμματική LR(k) πρέπει κατά τη ΣΑ αυτής να αρκεί η ανάγνωση μέχρι k προπορευόμενων συμβόλων (λεκτικών μονάδων) από την είσοδο για τον ακριβή καθορισμό κάθε κίνησης και κανόνα ελάττωσης
- Υπενθύμιση: Η ανάγνωση των προπορευόμενων λεκτικών μονάδων δεν τις καταναλώνει
- Συνήθως $k=1$

Λειτουργία Ακριβούς ΣΑ LR(1)

- Κατάσταση ΣΑ LR(1): Μία κατάσταση του ΣΑ LR(0), με την προσθήκη για κάθε στοιχείο του συνόλου των λεκτικών μονάδων που επιτρέπεται να ακολουθούν το αριστερό μέλος του *στη συγκεκριμένη κατάσταση*
- Η λειτουργία είναι παρόμοια με τη λειτουργία του ΣΑ LR(0), με τη διαφορά ότι ο πίνακας ACTION είναι δισδιάστατος, με τις στήλες να αντιστοιχούν στα προπορευόμενα σύμβολα

Προπορευόμενα Σύμβολα στον LR(1)

- Εκκινώντας με το στοιχείο $S' \rightarrow \bullet S \{ \text{EOF} \}$
- Βρες τις καταστάσεις LR(1) όπως νωρίτερα, με τη διαφορά ότι:

– Για ένα στοιχείο $\Sigma \equiv A \rightarrow a \bullet B \gamma$ X (όπου X σύνολο προπορευόμενων συμβόλων) και κάθε κανόνα $B \rightarrow \beta$ της γραμματικής:

$$\text{CLOSURE}(\Sigma) = \{ \Sigma \} \cup \{ B \rightarrow \bullet \beta Y \}$$

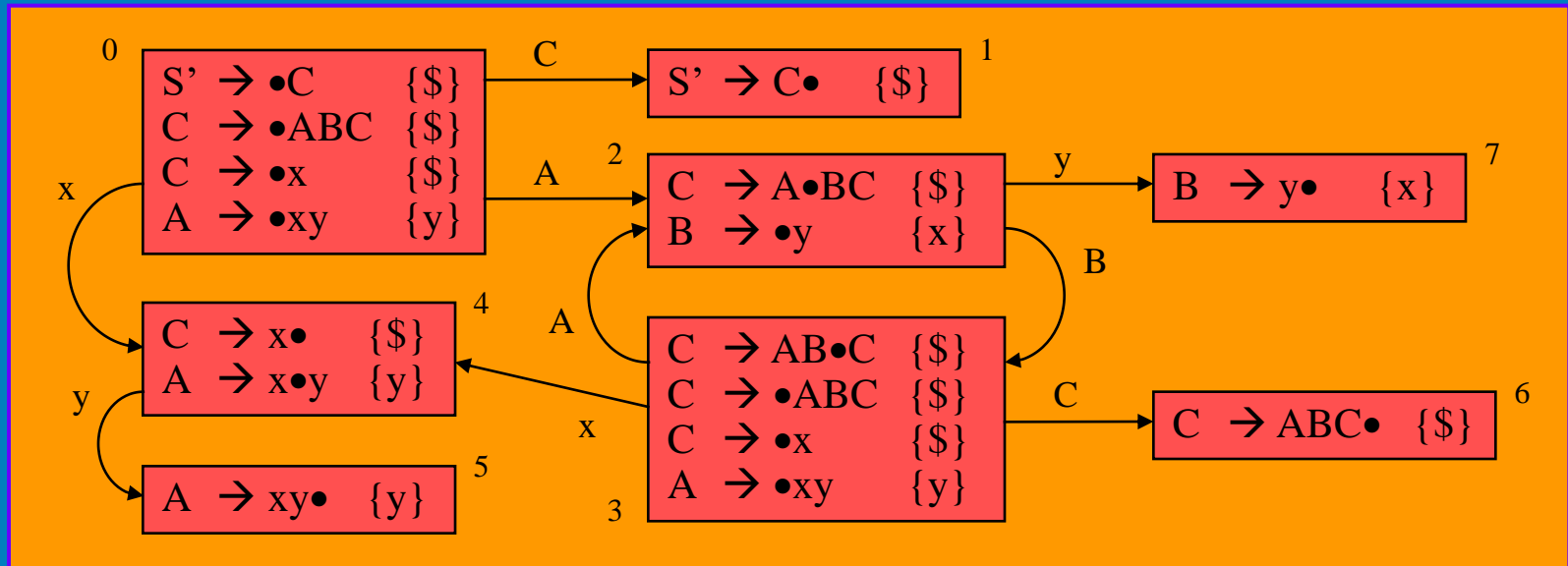
όπου:

$$Y = \text{FIRST}(\gamma) \text{ αν } \epsilon \notin \text{FIRST}(\gamma), \text{ ή}$$

$$Y = (\text{FIRST}(\gamma) - \{ \epsilon \}) \cup X \text{ αν } \epsilon \in \text{FIRST}(\gamma)$$

Παράδειγμα

- Για την προηγούμενη διευρυμένη γραμματική:



Πίνακες ΣΑ LR(1)

- Πίνακας ενέργειας ACTION
 - Αν $A \rightarrow \alpha \bullet a\beta$ $X \in I_i$ τότε $\text{ACTION}(i,a) =$ ολίσθηση
 - Αν $A \rightarrow \alpha \bullet X \in I_i$ και $A \neq S'$ τότε $\text{ACTION}(i,x) =$ ελάττωση με τον κανόνα $A \rightarrow \alpha$, για κάθε $x \in X$
 - Αν $S' \rightarrow \alpha \bullet \{\text{EOF}\} \in I_i$ τότε $\text{ACTION}(i,\text{EOF}) =$ αποδοχή
- Πίνακας επόμενης κατάστασης NEXT
 - Για κάθε σύμβολο x της γραμματικής:
Αν $\text{GOTO}(I_i,x) = I_j$ τότε $\text{NEXT}(i,x) = j$

Παράδειγμα

- Οι πίνακες LR(1) διαμορφώνονται ως εξής:

| ACTION / NEXT | | | | | | |
|---------------|------------|------------|------------|---|---|---|
| | x | y | EOF | A | B | C |
| 0 | ολίσθηση 4 | | | 2 | | 1 |
| 1 | | | αποδοχή | | | |
| 2 | | ολίσθηση 7 | | | 3 | |
| 3 | ολίσθηση 4 | | | 2 | | 6 |
| 4 | | ολίσθηση 5 | ελάττωση 2 | | | |
| 5 | | ελάττωση 3 | | | | |
| 6 | | | ελάττωση 1 | | | |
| 7 | ελάττωση 4 | | | | | |

- Δεν υπάρχει σύγκρουση!

Παράδειγμα

- Με είσοδο “xyyxyyx” ο ΣΑ λειτουργεί ως εξής:

| βήμα | στοίβα | είσοδος | κίνηση |
|------|-------------|-----------|-------------------------------------|
| 0 | 0 | xyyxyyx\$ | ολίσθηση 4 |
| 1 | 0x4 | yxyyx\$ | ολίσθηση 5 |
| 2 | 0x4y5 | yxyyx\$ | ελάττωση με τον κανόνα 3 (κόμβος 1) |
| 3 | 0A2 | yxyyx\$ | ολίσθηση 7 |
| 4 | 0A2y7 | xyyx\$ | ελάττωση με τον κανόνα 4 (κόμβος 2) |
| 5 | 0A2B3 | xyyx\$ | ολίσθηση 4 |
| 6 | 0A2B3x4 | yxx\$ | ολίσθηση 5 |
| 7 | 0A2B3x4y5 | yx\$ | ελάττωση με τον κανόνα 3 (κόμβος 3) |
| 8 | 0A2B3A2 | yx\$ | ολίσθηση 7 |
| 9 | 0A2B3A2y7 | x\$ | ελάττωση με τον κανόνα 4 (κόμβος 4) |
| 10 | 0A2B3A2B3 | x\$ | ολίσθηση 4 |
| 11 | 0A2B3A2B3x4 | \$ | ελάττωση με τον κανόνα 2 (κόμβος 5) |
| 12 | 0A2B3A2B3C6 | \$ | ελάττωση με τον κανόνα 1 (κόμβος 6) |
| 13 | 0A2B3C6 | \$ | ελάττωση με τον κανόνα 1 (κόμβος 7) |
| 14 | 0C1 | \$ | αποδοχή |

Προσέγγιση ΣΑ LR(1)

-
-
- Ο ΣΑ LR(1) έχει μεγάλο πλήθος καταστάσεων
 - Εκθετική αύξηση με τον αριθμό διαφορετικών προπορευόμενων του ίδιου συμβόλου
 - Υπερβολικά μεγάλοι πίνακες για ρεαλιστικές γλώσσες
- Υλοποίηση προσεγγίσεων του LR(1)
 - SLR(1)
 - LALR(1)

ΣΑ SLR(1)

- Στον SLR(1) προσεγγίζουμε τα σύνολα των λεκτικών μονάδων που μπορούν να ακολουθούν το αριστερό μέλος κάποιου στοιχείου με το αντίστοιχο σύνολο FOLLOW
- Με τον τρόπο αυτό δεν προκύπτουν πολλαπλές καταστάσεις που να διαφέρουν μεταξύ τους μόνο στα προπορευόμενα σύμβολα
- Η ανάλυση ανάγεται σε LR(0), όπου όμως τα σύνολα FOLLOW θέτουν περιορισμό για τις ελαττώσεις, μειώνοντας τις συγκρούσεις του LR(0)

ΣΑ LALR(1)

-
-
- Στον LALR(1) απλά δεν δημιουργούμε καταστάσεις που να διαφέρουν μόνο στα προπορευόμενα σύμβολα
 - Είτε προσθέτουμε προπορευόμενα σύμβολα σε ήδη υπάρχουσες καταστάσεις – και επαναλαμβάνουμε μέρος της ανάλυσης
 - Είτε κατασκευάζουμε αρχικά τις καταστάσεις LR(1) και μετά ενοποιούμε όσες διαφέρουν μόνο στα προπορευόμενα σύμβολα, ενώνοντας τα αντίστοιχα σύνολα προπορευόμενων συμβόλων

Σχέση μεταξύ ΣΑ LR

- $LR(0) \subset SLR(1) \subset LALR(1) \subset LR(1)$
 - Αν μια γραμματική ανήκει σε κάποια κατηγορία, θα ανήκει και στις μεγαλύτερες
 - πχ μια γραμματική $SLR(1)$ είναι και $LR(1)$
 - Αν μια γραμματική δεν ανήκει σε κάποια κατηγορία, δεν θα ανήκει ούτε στις μικρότερες
 - πχ μια γραμματική που δεν είναι $LR(1)$ δε μπορεί να είναι ούτε $LALR(1)$ ούτε $SLR(1)$ ούτε $LR(0)$

ΣΑ LALR(1) με το Yacc / Bison

- Πρώτο μέρος:
 - Κώδικας (πχ δηλώσεις C)
 - Δηλώσεις λεκτικών μονάδων, τελεστών και σημασιολογικών τύπων
- Δεύτερο μέρος:
 - Κανόνες παραγωγής
 - Κώδικας σημασιολογικής ανάλυσης που εκτελείται σε στοιχεία της γραμματικής
- Τρίτο μέρος:
 - Συμπληρωματικός κώδικας (πχ ορισμός main())

Γραμματικές LR και LL

-
-
- Οι ΣΑ γραμματικών LR είναι ισχυρότεροι των ΣΑ γραμματικών LL:
 - Χειρίζονται εύκολα γραμματικές που προκαλούν τις πιο συνηθισμένες συγκρούσεις FIRST/FIRST και FIRST/ FOLLOW στους δεύτερους!
- Επιτρέπουν αριστερή αναδρομή
- Επιτρέπουν κοινό αριστερό παράγοντα

Επίλυση Συγκρούσεων στη ΣΑ

- Σε κάθε περίπτωση σύγκρουσης, μπορούμε να προγραμματίσουμε το ΣΑ να προεπιλέγει κίνηση
 - ΣΑ γραμματικών που δεν είναι LL(1) ή LR(1)
 - Αλλάζει η γλώσσα της γραμματικής;
- Στο Yacc/Bison
 - Σε κάθε σύγκρουση ολίσθησης/ελάττωσης προτιμάται η ολίσθηση
 - Σε κάθε σύγκρουση ελάττωσης/ελάττωσης επιλέγεται ο πρώτος κανόνας

ΣΑ Γραμματικών Τελεστών

-
-
- Γραμματικές τελεστών είναι αυτές στις οποίες μεταξύ δύο μη τερματικών συμβόλων υπάρχει πάντα κάποιο τερματικό σύμβολο (τελεστής)
- Στη ΣΑ γραμματικών τελεστών, μπορούμε να προεπιλέγουμε κίνηση, ώστε να δίνουμε
 - προτεραιότητα σε κάποιους τελεστές έναντι άλλων
 - προσηταιριστικότητα (δεξιά, αριστερή ή καμία) σε τελεστές ίδιας προτεραιότητας

ΣΑ Γραμματικών Τελεστών

- Με την προεπιλογή:
 - Ενσωματώνουμε σημασιολογικές ιδιότητες της γλώσσας στη σύνταξή της
 - Απαλείφουμε τη γραμματική από ανεπιθύμητες ιδιότητες, όπως το διαφορούμενο!

Παράδειγμα

- Αν τ_1, τ_2 είναι τελεστές, τότε σε σύγκρουση

$$E \rightarrow E \tau_1 E \bullet \quad E \rightarrow E \bullet \tau_2 E$$

επιλέγουμε:

- ολίσθηση, αν ο τ_2 έχει υψηλότερη προτεραιότητα ή αν οι τ_1 και τ_2 έχουν ίση προτεραιότητα και δεξιά προσηταιριστικότητα,
- ελάττωση, αν ο τ_1 έχει υψηλότερη προτεραιότητα ή αν οι τ_1 και τ_2 έχουν ίση προτεραιότητα και αριστερή προσηταιριστικότητα,
- σφάλμα, σε κάθε άλλη περίπτωση