

# Μεταγλωττιστές

Γιώργος Δημητρίου

Μάθημα 1<sup>ο</sup>

# Γλώσσα - Μετάφραση

- Γλώσσα προγραμματισμού = Αναπαράσταση αλγορίθμων
  - Ευκολία χρήσης
  - Ακρίβεια και πληρότητα περιγραφής, όχι διφορούμενη!
  - Μία περιγραφή για όλες τις μηχανές
- Μετάφραση από την αλγοριθμική περιγραφή
  - Αρχική γλώσσα → Τελική γλώσσα
  - Μεταγλωττιστές & Διερμηνείς

# Εφαρμογές Μεταγλωττιστών

- Γλώσσες υψηλού επιπέδου
- Γλώσσες επεξεργασίας κειμένου / αναπαράστασης γραφικών
- Γλώσσες εκτύπωσης
- Γλώσσες περιγραφής υλικού
- Μερική χρήση μεταγλωττιστών
  - Επεξεργασία αρχείων
  - Ανίχνευση ιών

# Αναπαράσταση Γλώσσας

- Ακολουθία λεκτικών μονάδων  
 $( X ( I = ) A ( K / 2 ) + + MQ1$
- Συντακτικοί κανόνες γλώσσας  
 $assignment \rightarrow ID \text{ paren } = \text{ expression}$   
 $paren \rightarrow ( \text{ expression } ) | \epsilon$   
 $expression \rightarrow \text{ expression } ADDOP \text{ term } | \text{ term}$   
 $term \rightarrow \text{ term } MULOP \text{ ICONST } | ID \text{ paren}$
- Σημασιολογία γλώσσας  
 $X ( I ) = A ( K / 2 ) + MQ1$

# Λεκτικές Μονάδες

- Οι στοιχειώδεις λέξεις της γλώσσας  
( X ( I = ) A ( K / 2 ) + + M Q 1
  - Τα αναγνωριστικά είναι λεκτικές μονάδες
  - Οι σταθερές είναι λεκτικές μονάδες
  - Τα σύμβολα τελεστών και οι παρενθέσεις είναι λεκτικές μονάδες
  - Οι λέξεις-κλειδιά είναι λεκτικές μονάδες
  - Τα κενά δεν είναι λεκτικές μονάδες
  - Τα σχόλια δεν είναι λεκτικές μονάδες

# Γραμματική Γλώσσας

- Η γραμματική μιας γλώσσας αποτελείται από σύμβολα και κανόνες
- Τερματικά και μη τερματικά σύμβολα
  - Λεκτικές μονάδες: τερματικά σύμβολα
  - Αρχικό σύμβολο γραμματικής
- Κανόνες παραγωγής
$$\text{paren} \rightarrow ( \text{expression} ) \mid \varepsilon$$
  - paren: μη τερματικό σύμβολο
  - το δεξί μέρος ενός κανόνα είναι μια ακολουθία συμβόλων, ίσως η κενή ( $\varepsilon$ )

# Σημασιολογία Γλώσσας

- Σύνολο πληροφοριών που δεν προκύπτουν άμεσα από τη γραμματική της γλώσσας

$$X(I) = A(K/2) + MQ1$$

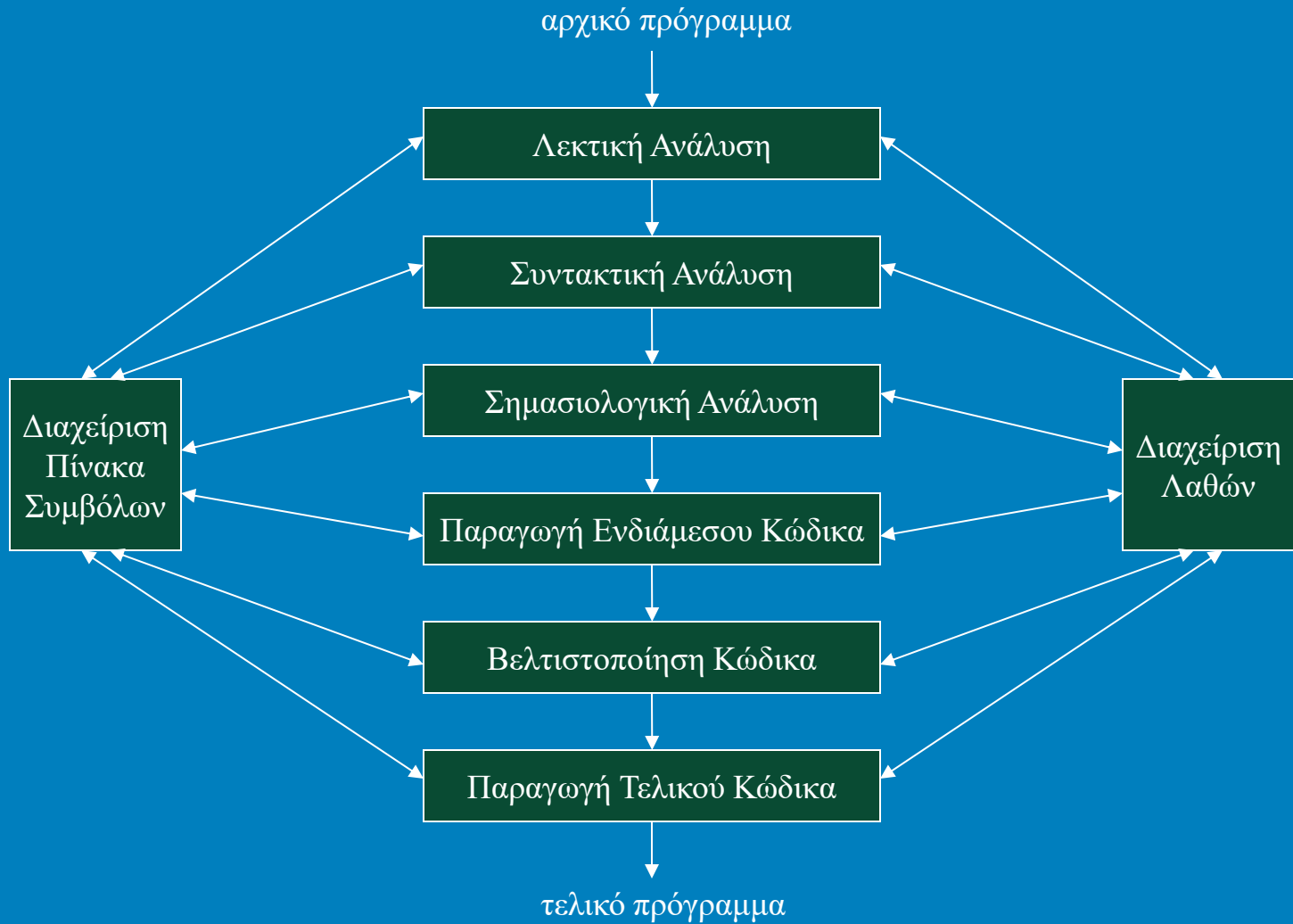
- Ονόματα και τύποι μεταβλητών
- Ακριβείς τελεστές ADDOP και MULOP
- $X$  και  $A$  πίνακες ή συναρτήσεις
- Μη ακέραιος τύπος δείκτη
- Τύποι παραστάσεων
- Συμβατότητα τύπων ανάθεσης
- Τιμή σταθεράς "2"

# Από μια Γλώσσα σε μια Άλλη

- Ισοδύναμα προγράμματα
  - Δύο προγράμματα είναι ισοδύναμα, εάν για τα ίδια δεδομένα εισόδου παράγουν τα ίδια δεδομένα εξόδου, για όλα τα δυνατά δεδομένα εισόδου
- Ο μεταγλωττιστής μεταφράζει ένα αρχικό πρόγραμμα (στην αρχική γλώσσα) σε ένα τελικό ισοδύναμο του πρώτου (στην τελική γλώσσα)
  - Ανάλυση του αρχικού - Σύνθεση του τελικού προγράμματος



# Φάσεις Μετάφρασης



# Λεκτική Ανάλυση

( X ( I = ) A ( K / 2 ) + + M Q 1

- Αναγνώριση λεκτικών μονάδων
  - Με χειρόγραφο επαναληπτικό πρόγραμμα
  - Με μηχανή καταστάσεων
  - Με τη βοήθεια μετα-μεταγλωττιστή
- Επιστροφή λεκτικών μονάδων και αρχικών σημασιολογικών τιμών (ονόματα, τελεστές, τιμές σταθερών)

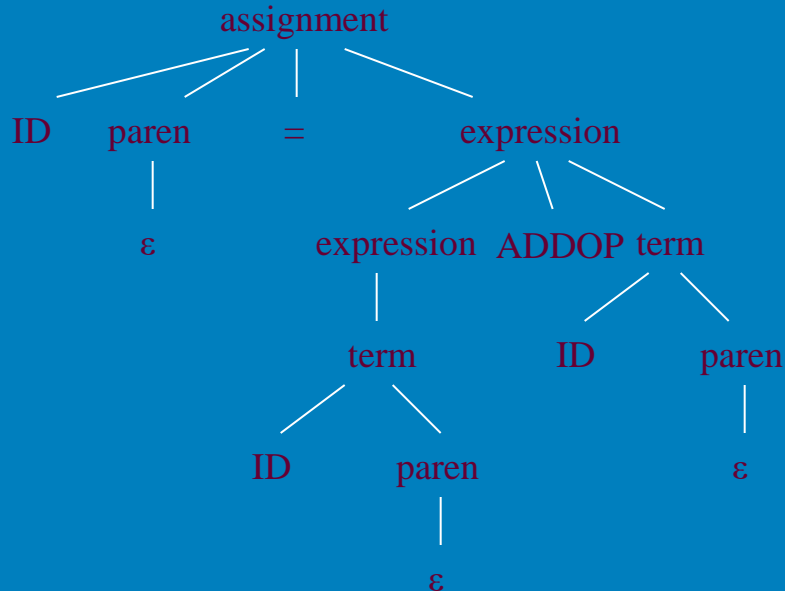
# Συντακτική Ανάλυση

- Διαδικασία συντακτικής ανάλυσης της ακολουθίας λεκτικών μονάδων (parsing)
- Από πάνω προς τα κάτω
  - Ξεκινά με το αρχικό σύμβολο “προβλέποντας” διαδοχικά νέες ακολουθίες συμβόλων.
- Από κάτω προς τα πάνω
  - Αναγνωρίζει σταδιακά συμβολοσειρές και “ελαττώνει” μέχρι το αρχικό σύμβολο.
- Αποδοχή ή απόρριψη προγράμματος

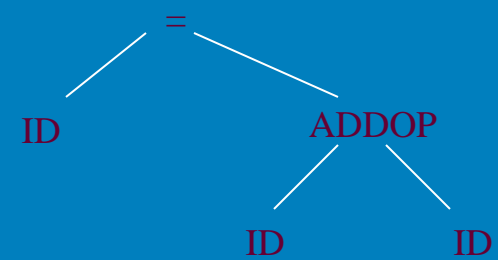
# Δέντρο Συντακτικής Ανάλυσης και Συντακτικό Δέντρο

ID = ID ADDOP ID

Δέντρο Συντακτικής Ανάλυσης



Συντακτικό Δέντρο



# Σημασιολογική Ανάλυση

- 
- 
- Συλλογή πληροφοριών σημασιολογίας
- Κατηγορικές γραμματικές
  - Επαύξηση συμβόλων με κατηγορήματα, πεδία που αποθηκεύουν ιδιότητες των συμβόλων
  - Τα κατηγορήματα μεταφέρουν τις τιμές τους από σύμβολο σε σύμβολο μέσω κατάλληλου κώδικα
    - κληρονομούμενα και συντιθέμενα

# Ενδιάμεσος Κώδικας

- 
- 
- Αφηρημένα συντακτικά δέντρα
  - Επαυξημένα συντακτικά δέντρα όπου δείχνεται η ροή ελέγχου και η ροή δεδομένων του αρχικού προγράμματος
- Τετράδες
  - / , K, 2, t1
  - deref, A, t1, t2
  - store, t2, -, X

# Βελτιστοποίηση Κώδικα

- Σταθερές και υπολογισμός σταθερών εκφράσεων
- Αναγνώριση άχρηστου κώδικα και κώδικα που ξαναχρησιμοποιείται (κοινές υποεκφράσεις)
- Βελτιστοποιήσεις σε βρόχους
  - Διαχείριση δεικτών επανάληψης
  - Μετακίνηση κώδικα εκτός βρόχου
- Παραλληλοποίηση κώδικα

# Παραγωγή Τελικού Κώδικα

- Επιλογή εντολών
- Δρομολόγηση εντολών
  - Εντολές φόρτωσης
  - Θέσεις καθυστέρησης διακλαδώσεων
- Δέσμευση καταχωρητών
- Τελευταίες βελτιστοποιήσεις
  - Μείωση αριθμού ακριβών εντολών
  - Δρομολόγηση βρόχων



# Πίνακας Συμβόλων

- Περιέχει όλα τα αναγνωριστικά που βρίσκονται στο αρχικό πρόγραμμα
- Τα αναγνωριστικά εισάγονται κατά τη λεκτική ή συντακτική ανάλυση
- Πληροφορίες συμπληρώνονται κατά τη συντακτική και τη σημασιολογική ανάλυση

# Διαχείριση Λαθών

- 
- 
- Από τις πιο βασικές λειτουργίες των μεταγλωττιστών
- Ανίχνευση λαθών
  - Συνήθως στη σύνταξη
- Ανάνηψη από λάθη
  - Συνέχιση μετάφρασης για ανίχνευση περισσότερων λαθών
  - Πόσο εύκολο είναι;

# Υλοποίηση Μεταγλωττιστή

- Παραγωγή σωστού κώδικα
- Παραγωγή γρήγορου κώδικα
- Γρήγορη μετάφραση
- Μικρές απαιτήσεις σε μνήμη
- Καλά διαγνωστικά μηνύματα
- Δυνατότητα ανάνηψης από λάθη
- Αξιοπιστία για κάθε αρχικό κώδικα
- Φορητότητα

# Περάσματα Μεταγλωττιστή

- Πέρασμα = Μια πλήρης επεξεργασία ολόκληρου του προγράμματος
- Ο αριθμός περασμάτων εξαρτάται από την πολυπλοκότητα της γλώσσας
- Συνήθως η παραγωγή ενδιάμεσου κώδικα γίνεται στο ίδιο πέραςμα με τη συντακτική και τη σημασιολογική ανάλυση
- Το επίπεδο βελτιστοποίησης επίσης επηρεάζει τον αριθμό περασμάτων

# Βοηθητικά Προγράμματα

- 
- 
- Προ-επεξεργαστές
- Συμβολομεταφραστές
- Φορτωτές
- Βιβλιοθήκες
- Μετα-μεταγλωττιστές
  - flex, bison

# Διερμηνεία

- 
- 
- Όχι τελικός κώδικας
- Οι εντολές του αρχικού προγράμματος εκτελούνται απ' ευθείας από την ενδιάμεση αναπαράσταση
  - Ουσιαστικά προσομοιώνεται η εκτέλεση του προγράμματος

# Γλώσσα Υλοποίησης

- Μπορεί ο μεταγλωττιστής να υλοποιηθεί στην αρχική γλώσσα της μετάφρασης;
  - Τεχνική εκκίνησης (bootstrapping):
    1. υλοποίηση σε τελική γλώσσα του μεταγλωττιστή ενός υποσυνόλου της αρχικής γλώσσας
    2. υλοποίηση σε αυτό το υποσύνολο του μεταγλωττιστή της συνολικής αρχικής γλώσσας
    3. τελική υλοποίηση του επιθυμητού μεταγλωττιστή