Μεταγλωττιστές

Εισαγωγή

Δημήτρης Μιχαήλ

Τμήμα Πληροφορικής και Τηλεματικής
Χαροκόπειο Πανεπιστήμιο

Βιβλιογραφία

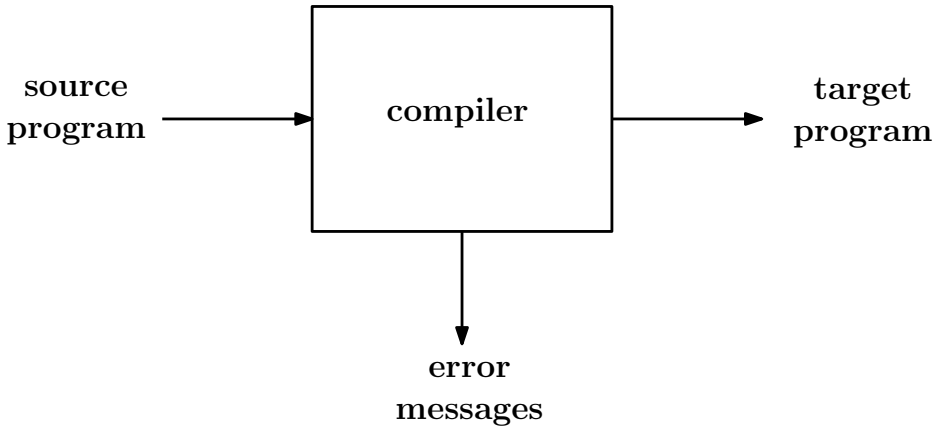
- Alfred V. Aho, Monica S. Lam, Ravi Sethi and Jeffrey D. Ullman. Compilers: Principles, Techniques, and Tools. 2nd edition. Addison-Wesley, 2007.
  http://dragonbook.stanford.edu/
- Νικόλαος Παπασπύρου και Εμμανουήλ Σκορδαλάκης. Μεταγλωττιστές, Εκδόσεις Συμμετρία.
- Κ. Λάζος, Π. Κατσαρός, Ζ. Καραΐσκος. Μεταγλωττιστές Γλωσσών Προγραμματισμού: θεωρία και πράξη. Εκδόσεις Θεσαλλονίκη 2004.
  http://delab.csd.auth.gr/~katsaros/CompilersBook.htm

Βιβλιογραφία

- Andrew W. Appel, Modern Compiler Implementation in Java. Cambridge University Press, 1998. http://www.cs.princeton.edu/~appel/modern
- Andrew W. Appel, Modern Compiler Implementation in C. Cambridge University Press, 1998. http://www.cs.princeton.edu/~appel/modern
- Steven S. Muchnick, Compiler Design and Implementation, Morgan Kaufmann Publishers, 1997.
- Allen I. Hollub, Compiler Design in C, Prentice Hall, 1990.

Μεταγλωττιστής

Είναι ένα πρόγραμμα που διαβάζει ένα πρόγραμμα γραμμένο σε μία γλώσσα και το μετατρέπει σε ένα πρόγραμμα γραμμένο σε άλλη γλώσσα.



Ο μεταγλωττιστής ενημερώνει τον χρήστη σχετικά με λάθη στο πρόγραμμα εισόδου.

Αριθμός Γλωσσών Προγραμματισμού

Πολύ μεγάλος αριθμός γλωσσών εισόδου

- C/C++
- Java
- Python
- Ruby
- Scala
- ...

Αντίστοιχα και ο αριθμός των δυνατών γλωσσών εξόδου είναι πολύ μεγάλος

- άλλη γλώσσα προγραμματισμού
- γλώσσα μηχανής ενός μικροεπεξεργαστή μέχρι και ενός υπερ-υπολογιστή

Έχοντας κατηγοριοποιήσει αντίστοιχα τις τεχνικές μπορούμε αυτή την στιγμή να υποστηρίξουμε όλες αυτές τις γλώσσες.

Χρόνος Ανάπτυξης Μεταγλωττιστή

- Το 1950 η συγγραφή ενός μεταγλωττιστή θεωρούταν ένα πολύ δύσκολο εγχείρημα.
- Ο πρώτος μεταγλωττιστής Fortran χρειάστηκε 18 ανθρωπο-έτη για να υλοποιηθεί.

Πλέον έχοντας συστηματοποιήσει αυτή την διαδικασία και τις τεχνικές υλοποίησης ενός μεταγλωττιστή έχουμε μειώσει αυτόν τον χρόνο.

Έχοντας τα απαραίτητα εργαλεία, ένας απλός μεταγλωττιστής μπορεί να υλοποιηθεί ως μία φοιτητική εργασία στην διάρκεια ενός προπτυχιακού μαθήματος.

Εφαρμογές

- binary translation: μετατροπή εκτελέσιμου μεταξύ αρχιτεκτονικών
- hardware synthesis: περιγραφή υλικού με την χρήση γλωσσών περιγραφής υλικού, π.χ VHDL (Very high-speed integrated circuit Hardware Description Language)
- Databases: Ερωτήματα σε γλώσσα Structured Query Language (SQL) μεταγλωττίζονται σε χαμηλού επιπέδου εντολές για αναζήτηση σε βάσεις δεδομένων
- Προγραμματιστικά εργαλεία για έλεγχο λαθών

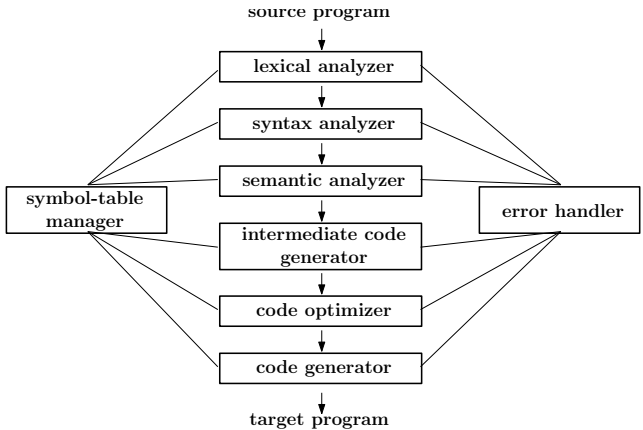
Μοντέλο Ανάλυσης-Σύνθεσης Μεταγλώττισης

Υπάρχουν δύο μέρη στην μεταγλώττιση

1. Ανάλυση
   Χωρίζει το πηγαίο πρόγραμμα σε μέρη και παρέχει μια ενδιάμεση αναπαράσταση του.
2. Σύνθεση
   Κατά την σύνθεση παράγεται το αποτέλεσμα από την ενδιάμεση αναπαράσταση.

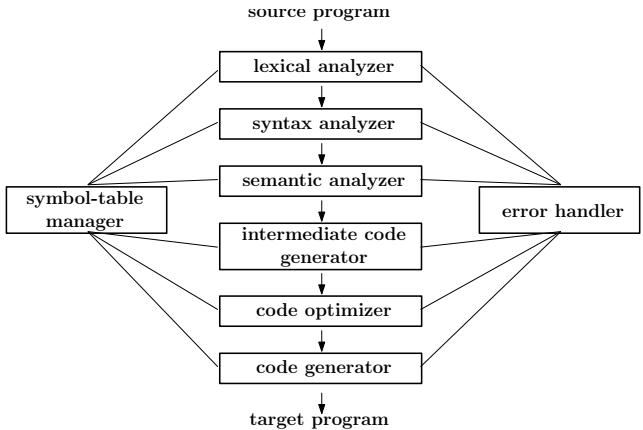
Οι φάσεις ενός μεταγλωττιστή

Ένας μεταγλωττιστής λειτουργεί σε φάσεις, κάθε μια από τις οποίες μετασχηματίζει το αρχικό πρόγραμμα από μία αναπαράσταση σε μια άλλη.



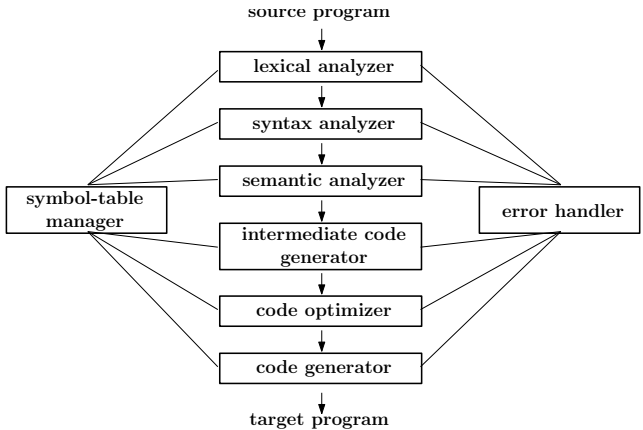
Οι φάσεις ενός μεταγλωττιστή

Οι πρώτες 3 φάσεις αποτελούν την ανάλυση.



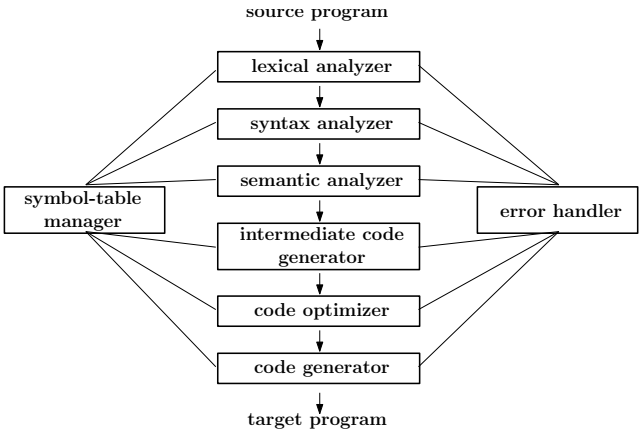
Οι φάσεις ενός μεταγλωττιστή

Οι υπόλοιπες 3 φάσεις αποτελούν την σύνθεση.



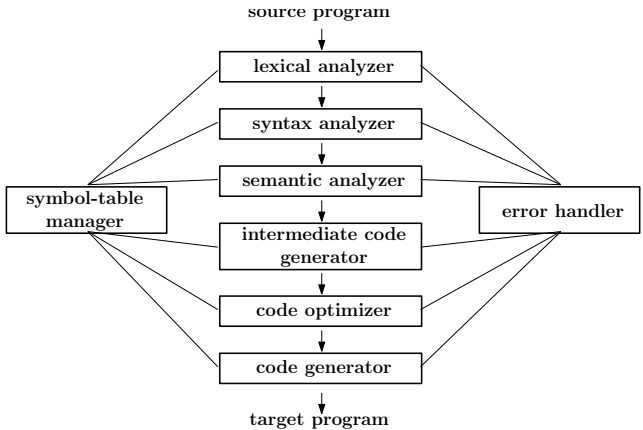
Οι φάσεις ενός μεταγλωττιστή

Καθόλη την διάρκεια της μεταγλώττισης υπάρχει επικοινωνία με τον πίνακα συμβόλων, ο οποίος αποθηκεύει πληροφορίες σχετικά με τα διάφορα αναγνωριστικά (identifiers) του προγράμματος, κ.τ.λ.



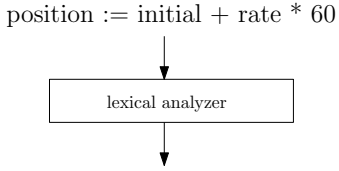
Οι φάσεις ενός μεταγλωττιστή

Αντίστοιχα η διαχείριση λαθών είναι μια διαδικασία που γίνεται σε όλες τις φάσεις με σκοπό την αναφορά αλλά και την δυνατότητα συνέχισης της μεταγλώττισης.



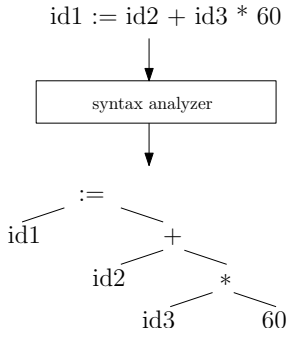
Οι φάσεις ενός μεταγλωττιστή

παράδειγμα



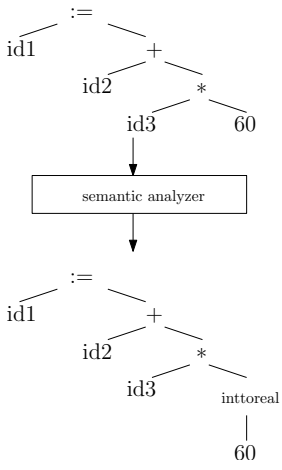
Οι φάσεις ενός μεταγλωττιστή

παράδειγμα



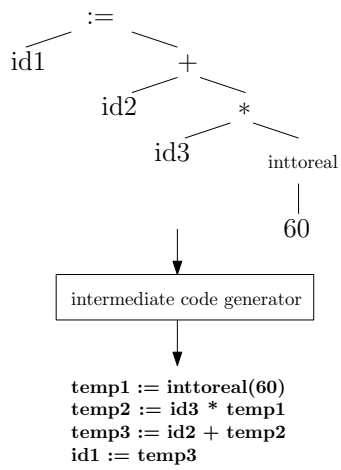
Οι φάσεις ενός μεταγλωττιστή

παράδειγμα



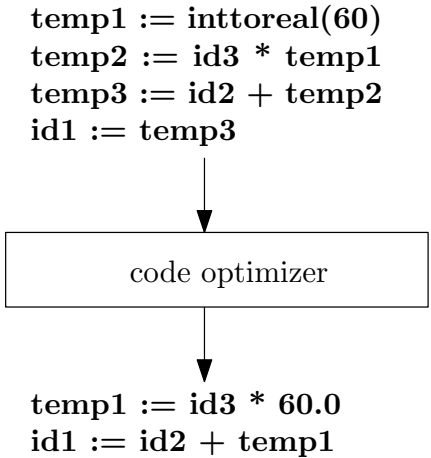
Οι φάσεις ενός μεταγλωττιστή

παράδειγμα



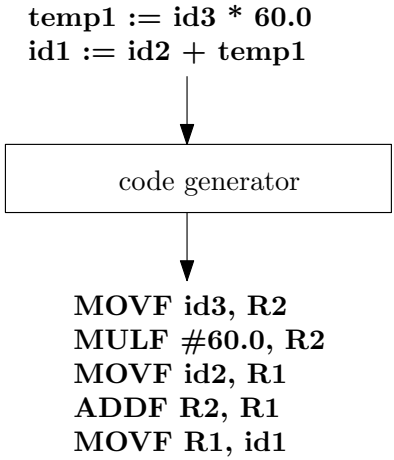
Οι φάσεις ενός μεταγλωττιστή

παράδειγμα



Οι φάσεις ενός μεταγλωττιστή

παράδειγμα



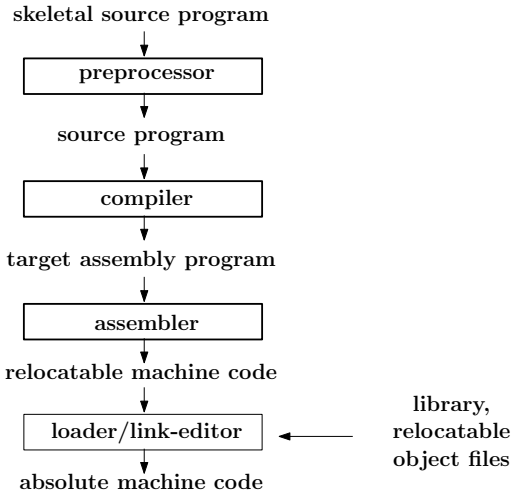
Το περιβάλλον ενός μεταγλωττιστή

Για την παραγωγή εκτελέσιμου κώδικα, εκτός του μεταγλωττιστή, είναι απαραίτητη η συμβολή και άλλων προγραμμάτων.

π.χ

- ένα πρόγραμμα είναι συνήθως χωρισμένο σε πολλά αρχεία
- ένα άλλο πρόγραμμα που λέγεται προεπεξεργαστής (preprocessor) είναι υπεύθυνο ώστε να συλλέξει τα διάφορα αρχεία

Το περιβάλλον ενός μεταγλωττιστή



Προεπεξεργαστής

Παράγει είσοδο στον μεταγλωττιστή.

Μπορεί να παρέχει τις εξής λειτουργίες:

1. **Macro processing**. Επιτρέπει στον χρήστη να ορίσει συντομογραφίες για μεγαλύτερες εκφράσεις.
2. **File inclusion**. Μας επιτρέπει να χρησιμοποιούμε αρχεία επικεφαλίδας.

Πολλές φορές μπορούμε να προσθέσουμε χαρακτηριστικά σε μια γλώσσα μέσω του preprocessor. Για παράδειγμα στις αρχικές υλοποιήσεις της γλώσσας C++ η υποστήριξη των templates γινόταν με την χρήση του preprocessor.

Assembler

Μερικοί μεταγλωττιστές παράγουν κώδικα assembly που μέσω ενός assembler μετατρέπεται σε κώδικα μηχανής.

Άλλοι μεταγλωττιστές παράγουν απευθείας μετατοπίσιμο (relocatable) κώδικα μηχανής που μπορεί να δωθεί απευθείας στο πρόγραμμα σύνδεσης (loader/link-editor).

Assembly Code

Ο κώδικας assembly είναι στην ουσία κώδικας μηχανής όπου δίνουμε

- ονόματα στις λειτουργίες (αντί για opcodes)
- ονόματα στις διευθύνσεις μνήμης.

π.χ

```
MOV a, R1
ADD #2, R1
MOV R1, b
```

είναι ισοδύναμο με

```
b = a + 2;
```

Two-Pass Assembly

Ένας απλός assembler κάνει δύο περάσματα πάνω από τον κώδικα.

1. Στο πρώτο πέρασμα διαβάζει τις εντολές και δημιουργεί έναν πίνακα συμβόλων για κάθε αναγνωριστικό που χρειάζεται στην μνήμη.
2. Στο δεύτερο πέρασμα μετατρέπει όλες τις εντολές σε κώδικα μηχανής.

Καθώς γράφει τον κώδικα μηχανής φροντίζει να προσθέσει και την πληροφορία σχετικά με τις εντολές που είναι μετατοπίσιμες στην μνήμη.

Μετατοπίσιμες εντολές είναι αυτές που αναφέρονται σε διευθύνσεις μνήμης που είναι σχετικές μετατοπίσεις της διεύθυνσης L που θα φορτωθεί το πρόγραμμα την ώρα της εκτέλεσης του.

Assembly

Σύνδεσμοι για περαιτέρω πληροφορίες για προγραμματιστές:

- http://www.faqs.org/docs/Linux-HOWTO/Assembly-HOWTO.html
- http://asm.sourceforge.net/
- http://sources.redhat.com/binutils/
- http://www.nasm.us/

Loader/Linker

Η φόρτωση ενός προγράμματος γίνεται αλλάζοντας όλες τις μετατοπίσιμες διευθύνσεις ανάλογα με την διεύθυνση όπου βρίσκεται το πρόγραμμα.

Το πρόγραμμα σύνδεσης είναι υπεύθυνο για να μπορούμε να δημιουργήσουμε ένα πρόγραμμα από πολλά αρχεία που περιέχουν μετατοπίσιμο κώδικα μηχανής.

Front-end & Back-end

Συνήθως χωρίζουμε έναν μεταγλωττιστή σε δύο μέρη, ένα που αφορά την γλώσσα και ένα που αφορά την παραγωγή τελικού κώδικα για την μηχανή.

Με αυτό τον τρόπο είναι πολύ εύκολο να φτιάξουμε μεταγλωττιστές και για άλλες μηχανές απλά αλλάζοντας μόνο το κομμάτι που αφορά το back-end.

Αριθμός Περασμάτων

Ένας μεταγλωττιστής μπορεί να λειτουργήσει σε ένα ή σε περισσότερα περάσματα της εισόδου. Θα δούμε τι ζητήματα προκύπτουν σε επόμενα μαθήματα.

Σε γενικές γραμμές μπορούμε να πούμε πως οι περισσότεροι σύγχρονοι μεταγλωττιστές είναι περισσότερων περασμάτων, αφού οι σύγχρονοι υπολογιστές έχουν άφθονη μνήμη.

Εργαλεία Κατασκευής Μεταγλωττιστών

Ο τομέας των μεταγλωττιστών είναι ένας τομέας όπου η θεωρία μας έχει δώσει πολύ σημαντικά εργαλεία κατασκευής μεταγλωττιστών.

Θα δούμε τέτοια εργαλεία κατά την διάρκεια του εξαμήνου καθώς θα υλοποιήσουμε έναν μεταγλωττιστή για μια απλή γλώσσα προγραμματισμού.