

Πανεπιστήμιο Θεσσαλίας

Τμήμα Πληροφορικής

Μεταγλωττιστές

Εργασία Εξαμήνου – Στάδιο 2

30 Νοεμβρίου 2018
(παράδοση 11 Ιανουαρίου)

Να κατασκευάσετε το Σημασιολογικό Αναλυτή για τη γλώσσα της εργασίας σας, μέσα από το Συντακτικό Αναλυτή (ΣΑ) που θα πρέπει να έχετε ήδη κατασκευάσει. Ο συνολικός Συντακτικός και Σημασιολογικός Αναλυτής (ΣΣΑ) πρέπει να καλείται μια φορά για δοθέν πρόγραμμα, και παράλληλα με τη συντακτική ανάλυση, θα πρέπει:

1. Να αποτιμά τα κατηγορήματα των συμβόλων της γραμματικής. Με βάση την περιγραφή της γλώσσας, να ορίσετε ένα σύνολο κατηγορημάτων για κάθε σύμβολο, έτσι ώστε να συγκεντρώνετε όποια πληροφορία χρειάζεστε για τη συνέχεια της μετάφρασης. Σχεδιάστε έτσι τον αναλυτή σας, ώστε να μπορείτε να προσθέτετε κατηγορήματα όταν διαπιστώνετε ότι είναι απαραίτητα. Μπορείτε να χρησιμοποιείτε καθολικές μεταβλητές για αποθήκευση πληροφορίας, προσπαθήστε όμως να χρησιμοποιείτε κατηγορήματα όσο μπορείτε, επειδή αυτά εισάγονται στη στοίβα αποτίμησης του ΣΑ κι επομένως διατηρούν την τιμή τους σε περιπτώσεις πολλαπλών εμφανίσεων του ίδιου συμβόλου ή σε περιπτώσεις φωλιασμένων δομών της γλώσσας.
2. Να προσθέτετε στον ΠΣ πληροφορίες για τα αναγνωριστικά που έχουν εισαχθεί από το ΛΑ ή το ΣΑ. Έτσι:
 - (α) Να κωδικοποιούνται στον ΠΣ οι τύποι και τα συνώνυμα τύπων, όπου αυτά υπάρχουν.
 - (β) Να προσδιορίζονται οι τύποι των μεταβλητών και των συναρτήσεων κατά τη δήλωσή τους.
 - (γ) Να αποδίδονται τιμές σε ονόματα σταθερών ή πιθανές αρχικές τιμές σε ονόματα μεταβλητών.
 - (δ) Να προσδιορίζονται ο αριθμός, οι τύποι, ο τρόπος περάσματος και τα ονόματα των παραμέτρων των υποπρογραμμάτων που δηλώνονται στο πρόγραμμα.
3. Αν η γλώσσα περιλαμβάνει εντολές *goto*, να τις συνδέει με τις αντίστοιχες ετικέτες. Για το σκοπό αυτό οι ετικέτες μπορούν να εισάγονται στον ΠΣ και να αντιμετωπίζονται σαν τα υπόλοιπα αναγνωριστικά.
4. Να κατασκευάζει το Αφηρημένο Συντακτικό Δέντρο (ΑΣΔ) του προγράμματος. Κάθε κόμβος κατασκευάζεται από τον κώδικα ενέργειας που συνοδεύει την αναγνώριση κάποιου συντακτικού κανόνα της γραμματικής. Ειδικότερα:
 - (α) Για τις εκφράσεις συνιστάται η χρήση τύπου δέντρου με μεταβλητό αριθμό παιδιών – ανάλογα με τον τελεστή κάθε κόμβου, ενώ για τις εντολές αρκεί η χρήση γραμμικής λίστας με δυνατότητα φωλιάσματος – για υποστήριξη σύνθετων και δομημένων εντολών. Για παράδειγμα, ο κόμβος μιας εντολής διακλάδωσης συνδέεται με τους κόμβους της προηγούμενης και της επόμενης εντολής, αλλά περιέχει και ένα ή δύο δείκτες σε κόμβους εντολών που αντιστοιχούν στις κατευθύνσεις της διακλάδωσης. Ας σημειωθεί ότι ο ίδιος κόμβος περιέχει και ένα δείκτη σε ρίζα δέντρου έκφρασης. Γενικά, οι περισσότεροι κόμβοι εντολών περιέχουν δείκτες σε δέντρα έκφρασης.
 - (β) Η δομή των κόμβων του ΑΣΔ πρέπει να είναι τέτοια, ώστε να περικλείει όλη την πληροφορία που παράγεται από τη σημασιολογική ανάλυση των ίδιων και είναι απαραίτητη για τη σημασιολογική ανάλυση των γειτονικών κόμβων. Έτσι, ένας κόμβος έκφρασης πρέπει να αποθηκεύει πληροφορία για τον τύπο του, η οποία παράγεται από τη σημασιολογική ανάλυσή του και χρησιμοποιείται για τη σημασιολογική ανάλυση των γονέων του.

(γ) Το δέντρο πρέπει να είναι όσο γίνεται πιο αφηρημένο, ώστε να είναι μικρό και να περιλαμβάνει μόνο τους αναγκαίους κόμβους για τη μετέπειτα ανάλυση του προγράμματος. Για παράδειγμα, κόμβοι για την αποθήκευση διαχωριστικών – πχ παρενθέσεων – είναι σκόπιμο να παραλείπονται, ενώ οι εκφράσεις είναι προτιμότερο να αποθηκεύονται έτσι, ώστε σε κάθε εσωτερικό κόμβο τους να αποθηκεύεται ο αντίστοιχος τελεστής, αντί κάποιου μη τερματικού συμβόλου της γραμματικής.

Να έχετε υπ' όψη ότι ο τύπος του κόμβου του ΑΣΔ θα παίζει σημαντικό ρόλο στη βελτιστοποίηση και την παραγωγή τελικού κώδικα, γι' αυτό θα πρέπει να είναι εύκολος στο χειρισμό και να μπορεί να τροποποιείται εύκολα. Καλό θα είναι να διαχωρίσετε τον κώδικα C που διαχειρίζεται το ΑΣΔ σε κάποιο αρχείο έξω από το κυρίως αρχείο του ΣΣΑ.

5. Να δεσμεύει θέσεις στο χώρο δεδομένων (ΧΔ) για τις μεταβλητές του προγράμματος, δημιουργώντας πίνακες δέσμευσης για κάθε εμβέλεια που μεταφράζεται. Για τη γλώσσα της εργασίας σας καθορίστε τις κατηγορίες αποθήκευσης που υποστηρίζονται, το δέσιμο των μεταβλητών, το φώλιασμα των εμβελιών, και συμπληρώστε στον ΠΣ τα στοιχεία που θεωρείτε απαραίτητα για την παραγωγή τελικού κώδικα της επίλυσης αναφοράς κάθε μεταβλητής.
6. Να εκτελεί τον έλεγχο των σημασιολογικών κανόνων της γλώσσας. Πιο συγκεκριμένα:
 - (α) Να ελέγχει την ορθότητα όλων των δηλώσεων. Έτσι για παράδειγμα δεν πρέπει να υπάρχουν πολλαπλές δηλώσεις για το ίδιο αναγνωριστικό στην ίδια εμβέλεια. Η ακόμα οι δηλώσεις συναρτήσεων πρέπει να ταιριάζουν με πιθανή προηγούμενη δήλωση των πρωτοτύπων τους, ο τύπος δήλωσης μιας σταθεράς πρέπει να ταιριάζει με την τιμή που της αποδίδεται, και ο τύπος μιας μεταβλητής πρέπει να ταιριάζει με την πιθανή αρχική της τιμή.
 - (β) Να ελέγχει τη χρήση όλων των αναγνωριστικών σε σχέση με τη δήλωσή τους. Η χρήση αυτών πρέπει να γίνεται με τρόπο συμβατό με την αντίστοιχη δήλωση και στην εμβέλεια στην οποία έχουν δηλωθεί.
 - (γ) Να ελέγχει τους τύπους των επιμέρους εκφράσεων σε μεγαλύτερες εκφράσεις. Αν αυτό κριθεί αναγκαίο, κόμβοι μετατροπής συμβατών τύπων πρέπει να εισάγονται στο ΑΣΔ.
 - (δ) Να ελέγχει τη σημασιολογική ορθότητα των εντολών του προγράμματος. Για παράδειγμα, εδώ ελέγχονται οι τύποι στις αναθέσεις και προστίθενται τελικοί κόμβοι μετατροπής τύπων από το δεξί στο αριστερό μέλος μιας ανάθεσης, αν αυτό είναι απαραίτητο. Επίσης ελέγχονται οι κλήσεις υποπρογραμμάτων για την ορθότητα του αριθμού και του τύπου των παραμέτρων. Η ακόμα ελέγχονται οι τύποι παραστάσεων που υπεισέρχονται σε συγκεκριμένες εντολές, όπως εντολές διακλαδώσεων και εντολές βρόχων. Τέλος, εδώ ελέγχονται τα άλματα εντολών *goto*, και γενικά κάθε είδους εντολές ελέγχου ροής σε καθεμιά από τις τρεις γλώσσες της εργασίας σας.
7. Σε περίπτωση σφάλματος συντακτικής ανάλυσης να σταματά την παραγωγή κώδικα. Σε περίπτωση σημασιολογικού σφάλματος, ο ΣΣΑ να τερματίζει άμεσα τη λειτουργία του.
8. Για τους σκοπούς της παρούσας άσκησης να εκτυπώνει με κατάλληλο τρόπο το ΑΣΔ και το ΧΔ στο τέλος κάθε μονάδας του προγράμματος εισόδου.

Η σημασιολογική ανάλυση γίνεται μέσα από τις ενέργειες του μετα-εργαλείου Bison. Ας σημειωθεί ότι επειδή το Bison υποστηρίζει ένα μόνο κατηγορημα ανά σύμβολο της γραμματικής, η υλοποίηση πολλαπλών κατηγορημάτων θα πρέπει να γίνει με το δομημένο τύπο `struct` της C. Τις πιο πολλές φορές, ένα σύμβολο της γραμματικής αποδίδει έναν κόμβο του ΑΣΔ, οπότε τα κατηγορήματα του συμβόλου μπορούν να ενσωματωθούν στη δομή του αντίστοιχου κόμβου.