

Πανεπιστήμιο Θεσσαλίας
Τμήμα Πληροφορικής και Τηλεπικοινωνιών

Μεταγλωττιστές

Εργασία Εξαμήνου – Στάδιο 1

Νοέμβριος 2022
(παράδοση πριν τις γιορτές)

Να κατασκευάσετε το Λεκτικό και το Συντακτικό Αναλυτή (ΛΑ, ΣΑ) για τη γλώσσα της εργασίας σας. Για κάποιο τυχαίο πρόγραμμα εισόδου, ο ΣΑ καλείται μία μοναδική φορά, για να αποφασίσει αν το πρόγραμμα είναι συντακτικά ορθό, ενώ ο ΛΑ καλείται επαναληπτικά από το ΣΑ, κάθε φορά που ο δεύτερος χρειάζεται να προχωρήσει σε επόμενη Λεκτική Μονάδα (ΛΜ) του προγράμματος.

Α. Πριν προχωρήσετε στην κατασκευή του ΛΑ, θα πρέπει να σχεδιάσετε το Διάγραμμα Μετάβασης (ΔΜ) της γλώσσας της εργασίας σας σε ενιαία μορφή. Τοποθετήστε πρώτα τις λέξεις-κλειδιά, στη συνέχεια προσθέστε τις σταθερές και τα αναγνωριστικά, και τέλος συμπληρώστε τους τελεστές και τα διαχωριστικά σύμβολα. Να κατασκευάσετε έναν πίνακα με όλες τις τελικές καταστάσεις του ΔΜ, όπου να σημειώσετε τις ΛΜ που επιστρέφονται, καθώς και την οπισθοδρόμηση που πρέπει να γίνεται σε κάθε τελική κατάσταση.

Προχωρήστε τώρα στην υλοποίηση του ΛΑ, ο οποίος θα πρέπει:

1. Να επιστρέφει τον κωδικό της ΛΜ που βρίσκει κάθε φορά που καλείται.
2. Να επιστρέφει τα ονόματα των αναγνωριστικών που συναντά, ή εναλλακτικά να εισάγει στον Πίνακα Συμβόλων (ΠΣ) τα αναγνωριστικά που συναντά, επιστρέφοντας το δείκτη στον ΠΣ που αντιστοιχεί στο τρέχον αναγνωριστικό.
3. Να επιστρέφει τις τιμές των σταθερών που συναντά. Για μη αριθμητικές σταθερές να επιστρέφει τη συμβολοσειρά της λέξης που αναγνώρισε.
4. Να χειρίζεται σφάλματα λεκτικής ανάλυσης με εκτύπωση μηνύματος που να περιλαμβάνει τον αριθμό γραμμής και την ίδια τη γραμμή του αρχείου εισόδου στην οποία σημειώθηκε το σφάλμα. Να προσπαθεί να συνεχίσει σβήνοντας το λανθασμένο χαρακτήρα, και αν αυτό δεν είναι δυνατό, τότε να τερματίζει τη λειτουργία του.
5. Για τους σκοπούς της παρούσας άσκησης να εκτυπώνει κατάλληλα όλες τις τιμές που επιστρέφει, πέρα από τυχόν μηνύματα λάθους.

Για την κατασκευή του ΛΑ μπορείτε να χρησιμοποιήσετε είτε τη μέθοδο του απ' ευθείας προγραμματισμού του ΔΜ που σχεδιάσατε, είτε το μετα-εργαλείο flex. Η κωδικοποίηση των ΛΜ να γίνει με το συμβολισμό που δίνεται στην περιγραφή της γλώσσας που μεταφράζετε. Για παράδειγμα, το σύμβολο `ID` να χρησιμοποιηθεί για την κωδικοποίηση των αναγνωριστικών. Σε περίπτωση που χρησιμοποιήσετε το flex, θα πρέπει οι πληροφορίες που ο ΛΑ επιστρέφει πέρα από τον κωδικό της ΛΜ, να επιστρέφονται μέσω της μεταβλητής `yylval`.

Για να γίνει δεκτή η παράδοση του ΛΑ αν χρησιμοποιήσετε το flex, θα πρέπει ο πιο πάνω πίνακας τελικών καταστάσεων του ΔΜ να συμπληρωθεί με την αντιστοίχιση κάθε αριθμού κατάστασης του δικού σας ΔΜ με τον αριθμό κατάστασης του ΔΜ που κατασκευάζει το flex. Για το σκοπό αυτό, θα πρέπει να ανοίξετε το αρχείο που παράγει το flex ("`lex.yy.c`") και να εντοπίσετε τις καταστάσεις του ΔΜ που αναγνωρίζουν τις ΛΜ της γλώσσας. Επειδή το αρχείο αυτό παράγεται αυτόματα, φροντίστε να το μετονομάσετε σε άλλο κατάλληλο όνομα, για να μην μπερδευτείτε και το χάσετε σε κάποια επόμενη εκτέλεση του flex. Το συγκεκριμένο αρχείο που χρησιμοποιήσατε θα πρέπει να παραδοθεί μαζί με το υπόλοιπο υλικό του ΛΑ.

B. Ο ΣΑ θα πρέπει:

1. Να αναγνωρίζει ή να απορρίπτει το πρόγραμμα με βάση τη γραμματική της γλώσσας. Πιο συγκεκριμένα: Για να κατασκευαστεί ένας ντετερμινιστικός ΣΑ για τη γλώσσα της εργασίας σας, πρέπει η γραμματική να μην είναι διφορούμενη. Για το σκοπό αυτό πρέπει να μετατρέψετε τη διφορούμενη γραμματική που σας δίνεται σε μη διφορούμενη με έναν από τους εξής δύο τρόπους: (α) Να μετασχηματίσετε τη γλώσσα σε μη διφορούμενη, ή (β) να χρησιμοποιήσετε την προτεραιότητα και την προσεταιριστικότητα των τελεστών της γλώσσας, ώστε η ανάλυση να οδηγεί σε μοναδικό συντακτικό δέντρο. Εναλλακτικά, για ορισμένους κανόνες, στους οποίους η σύγκρουση που εμφανίζεται λύνεται πάντα προς μία κατεύθυνση – όπως γίνεται στην περίπτωση του *ξεκρέμαστος else*, προτιμάμε να επιβάλουμε αυτή την κατεύθυνση στο ΣΑ.
2. Να χειρίζεται σφάλματα συντακτικής ανάλυσης με εκτύπωση κατάλληλων μηνυμάτων που να περιλαμβάνουν τον αριθμό γραμμής και την ίδια τη γραμμή του αρχείου εισόδου στην οποία σημειώθηκε κάθε σφάλμα. Ειδικότερα: Για ανάνηψη από κάποιο σφάλμα να χρησιμοποιηθεί η μέθοδος πανικού σε συνδυασμό με τη στρατηγική των κανόνων παραγωγής σφαλμάτων για τα πιο κοινά σφάλματα, όπως η απουσία διαχωριστικών. Μετά το πρώτο σφάλμα να συνεχίζεται η ανάλυση μέχρι έναν αριθμό 5 το πολύ σφαλμάτων.
3. Να εισάγει στον ΠΣ τα αναγνωριστικά του προγράμματος, αν δεν έχουν εισαχθεί από το ΛΑ. Ακόμα, θα πρέπει να διαχειρίζεται τις εμβέλειες του προγράμματος στον ΠΣ, σύμφωνα με την περιγραφή της γλώσσας που μεταφράζετε.
4. Για τους σκοπούς της παρούσας άσκησης να εκτυπώνει με κατάλληλο τρόπο το περιεχόμενο του ΠΣ στο τέλος κάθε εμβέλειας.

Η χρήση του μετα-εργαλείου *bison* είναι αναγκαία, τόσο για την κατασκευή του ΣΑ, όπου μπορούμε να χειριστούμε διφορούμενες γραμματικές χωρίς την ανάγκη μετασχηματισμού, όσο και για την ενσωμάτωση σε αυτόν του σημασιολογικού αναλυτή. Με το *bison* μπορούμε επίσης να υλοποιήσουμε ευκολότερα τη μέθοδο πανικού για ανάνηψη από συντακτικά σφάλματα.

Για να γίνει δεκτή η παράδοση του ΣΑ, θα πρέπει να μελετήσετε το αρχείο ανάλυσης της γραμματικής που παράγει το *bison* (με επιλογή “-v” προκύπτει το αρχείο με κατάληξη “.output”), και (α) σε περίπτωση που δεν μετασχηματίσετε τη γραμματική σας σε μη διφορούμενη, θα πρέπει να βρείτε και να υποδείξετε τις καταστάσεις στις οποίες εμφανίζονται συγκρούσεις, καθώς και τις καταστάσεις όπου οι συγκρούσεις του ΣΑ επιλύονται με βάση την προτεραιότητα και προσεταιριστικότητα των τελεστών, δίνοντας και τα εμπλεκόμενα στοιχεία του ΣΑ, ενώ (β) σε περίπτωση που μετασχηματίσετε τη γραμματική σας σε μη διφορούμενη, θα πρέπει να βρείτε και να υποδείξετε τις καταστάσεις στις οποίες εμφανίζονται συγκρούσεις *πριν το μετασχηματισμό*, δίνοντας και τα εμπλεκόμενα στοιχεία του ΣΑ, καθώς και τις νέες καταστάσεις που προκύπτουν *μετά το μετασχηματισμό*, όπου οι συγκρούσεις θα έχουν επιλυθεί.

Όσο αφορά τον ΠΣ, μπορείτε να τον σχεδιάσετε με όποια δομή θέλετε, με βαθμολόγηση όμως ανάλογη της ταχύτητας προσπέλασής του. Προς το παρόν η μόνη πληροφορία που αποθηκεύεται στον ΠΣ για κάποιο αναγνωριστικό είναι το όνομά του και το βάθος φωλιάσματος της εμβέλειάς του.

Για επαλήθευση τόσο του ΛΑ όσο και του ΣΑ χρησιμοποιήστε τα αρχεία εισόδου που επισυνάπτονται του αρχείου περιγραφής της γλώσσας της εργασίας σας.