

Εργασία Εξαμήνου

Μέρος Β' – Δημιουργία Συντακτικού Αναλυτή με τη γεννήτρια BISON

Εισαγωγή

Ο ρόλος του συντακτικού αναλυτή – ΣΑ (syntax analyzer) είναι να διαπιστώσει τη συντακτική ορθότητα των δομών του πηγαίου κώδικα βάσει της γραμματικής της γλώσσας. Η είσοδος στον συντακτικό αναλυτή - ΣΑ (parser) είναι μια ακολουθία από λεκτικές μονάδες (tokens), που παράγονται από τον λεκτικό αναλυτή - ΛΑ (lexical analyzer), μετά το διαχωρισμό και την αναγνώριση των λεξημάτων της συμβολοσειράς εισόδου. Οι λεκτικές μονάδες επιστρέφονται από τον ΛΑ **όχι μονομιάς** αλλά μετά από συνεχή αιτήματα του ΣΑ κατά τη διάρκεια των ελέγχων που πραγματοποιεί.

Για να μπορέσετε να προχωρήσετε στην ανάπτυξη του Β' μέρους της εργασίας θα πρέπει να έχετε αποκτήσει τις απαραίτητες θεωρητικές γνώσεις για την **περιγραφή γραμματικών κανόνων** με τη σημειογραφία **EBNF** και για τον τρόπο λειτουργίας των **συντακτικών και σημασιολογικών αναλυτών** που έχουν διδαχθεί στο θεωρητικό μέρος του μαθήματος και να βεβαιωθείτε ότι είστε σε θέση να περιγράψετε τους γραμματικούς κανόνες της γλώσσας του εργαστηρίου με τη σημειογραφία **EBNF**.

Για να βοηθηθείτε στη δημιουργία του συντακτικού αναλυτή μέσω της γεννήτριας BISON, μελετήστε τα έγγραφα [ΕΙΣΑΓΩΓΗ ΣΤΟ BISON.pdf](#) και [Εισαγωγή στο BISON - 2.pdf](#) που θα βρείτε στον κατάλογο » 3-ΕΡΓΑΣΤΗΡΙΟ » 3-BISON στα έγγραφα της πλατφόρμας Open eClass. Στη συνέχεια, για περισσότερες πληροφορίες μελετήστε το εγχειρίδιο [Οδηγός FLEX & BISON](#) που θα βρείτε στον ίδιο φάκελο.

Αφού ολοκληρώσετε την εργασία, θα πρέπει να είστε σε θέση να:

- σχεδιάζετε κανόνες παραγωγής με τη σημειογραφία EBNF για την περιγραφή της γραμματικής μιας γλώσσας προγραμματισμού
- γράφετε κώδικα bison για τη δημιουργία ανεξάρτητου συντακτικού αναλυτή με στοιχεία σημασιολογικής ανάλυσης
- προσαρμόζετε τους κώδικες flex και bison ώστε να συνεργάζονται μεταξύ τους για τη δημιουργία ολοκληρωμένου περιβάλλοντος ανάλυσης
- εμπλουτίζετε τη γραμματική μιας γλώσσας ώστε να αναγνωρίζει γνωστά προειδοποιητικά λεκτικά και σημασιολογικά λάθη

Σκοπός αυτής της εργασίας είναι να αποκτήσετε σχετική εμπειρία στην περιγραφή γραμματικών μέσω της σημειολογίας EBNF και ειδικότερα να εξοικειωθείτε με την ανάπτυξη συντακτικών. Για να αναπτύξετε τον συντακτικό σας αναλυτή θα χρησιμοποιήσετε τη γεννήτρια BISON ακολουθώντας τις αναρτημένες οδηγίες. Θα δοκιμάσουμε τους κώδικες που θα καταθέσετε χρησιμοποιώντας τον BISON που είναι εγκατεστημένος στο εργαστήριο, οπότε βεβαιωθείτε ότι οι λύσεις σας λειτουργούν σωστά στο περιβάλλον του εργαστηρίου (με πρόσβαση εξ αποστάσεως) πριν τις υποβάλλετε μέσω του εργαλείου των Εργασιών του Open eClass.

Πριν προχωρήσετε στα επί μέρους ζητούμενα της εργασίας, πρέπει να περιγραφούν, με τη βοήθεια της EBNF σημειογραφίας, οι εκφράσεις (expressions) της γλώσσας, να δημιουργηθούν δηλαδή οι κανόνες παραγωγής της γραμματικής της.

ΜΕΡΟΣ Β-1: Ανάπτυξη κώδικα bison για δημιουργία ανεξάρτητου ΣΑ

Στο μέρος αυτό θα χρειαστεί κατ' αρχάς να ολοκληρώσετε τον πρότυπο ημιτελή κώδικα simple-bison-code.y μέσα από το συμπιεσμένο αρχείο με όνομα simple-bison-code.zip που θα βρείτε μέσα στον φάκελο [ΕΚΦΩΝΗΣΕΙΣ ΕΡΓΑΣΙΩΝ](#) του φακέλου [4-Υλικό για Εργασίες](#). Ο πρότυπος κώδικας υλοποιεί τη συντακτική ανάλυση πηγαίου κώδικα της γλώσσας εργαστηρίου που περιέχει προσθέσεις ακέραιων αριθμών (δεκαδικού συστήματος) και μεταβλητών.

Το αρχείο που σας δίνεται, περιέχει έναν αριθμό από **"FILL ME"** τα οποία και πρέπει να αντικαταστήσετε με τον σωστό κώδικα bison προκειμένου να επιτύχετε την συντακτική ανάλυση κώδικα που περιέχει αναθέσεις απλών αριθμητικών εκφράσεων σε μεταβλητές. Αφού πραγματοποιήσετε όλες τις αντικαταστάσεις, δοκιμάστε την εκτέλεσή του αρχείου σας με τον συντακτικά ορθό πηγαίο κώδικα που θα εισάγετε μέσα στο αρχείο input.txt, συγκρίνετε τα αποτελέσματα της εκτέλεσης που θα βρίσκονται στο αρχείο output.txt με τα αναμενόμενα και προσπαθήστε να κατανοήσετε τη λειτουργία της γεννήτριας.

Αφού πετύχετε το παραπάνω ζητούμενο, **ολοκληρώσετε τον συντακτικό αναλυτή** σας ώστε να αναγνωρίζει όσες περισσότερες εκφράσεις της γλώσσας του εργαστηρίου μπορείτε. Φροντίστε επίσης απαραιτήτως να εισάγετε με ευανάγνωστο τρόπο τα δικά σας σχόλια στον κώδικα που θα δημιουργήσετε.

Για την περιγραφή της σύνταξης των εκφράσεων με την EBNF είναι αναγκαίο να γίνει σαφής ο όρος atom που αποτελεί συνήθως ένα από τα βασικότερα στοιχεία της δομής μιας έκφρασης. Τα atoms δεν αποτελούν λεκτικές μονάδες αλλά απλά μέρος της σύνταξης των εκφράσεων (ρητά δηλωμένες συμβολοσειρές), δηλαδή αναγνωρίζονται απευθείας από τον συντακτικό αναλυτή χωρίς αυτός να ζητήσει τη βοήθεια του λεκτικού αναλυτή, π.χ. στην έκφραση:

```
expr: ID '+' NUM
```

το expr θεωρείται μη-τερματικός όρος, τα ID και NUM tokens και το '+' θεωρείται atom.

Φροντίστε να ελέγξετε εξαντλητικά τον ΣΑ που θα δημιουργηθεί, χρησιμοποιώντας ως είσοδο έναν δικό σας πηγαίο κώδικα ως αρχείο εισόδου που θα περιέχει και σωστές αλλά και λανθασμένες εκφράσεις της γλώσσας. Όταν μια έκφραση αναγνωρίζεται να εμφανίζεται στο αρχείο εξόδου αριθμημένη.

Αναρτήστε έγκαιρα τον ολοκληρωμένο κώδικά σας, σωστά σχολιασμένο για να γίνεται κατανοητός, μαζί με τα αποτελέσματα των ελέγχων σας. Η κατάθεσή σας θα είναι ένα συμπιεσμένο αρχείο ZIP με όνομα που θα αρχίζει με τον κωδικό της ομάδας υποχρεωτικά και που θα περιλαμβάνει τα παρακάτω:

1. Το αρχείο για τη δημιουργία του συντακτικού αναλυτή (ολοκληρωμένος κώδικας bison)
2. Συνοδευτικά αρχεία εισόδου - εξόδου (input.txt - output.txt) που θα περιλαμβάνουν αρκετές περιπτώσεις αναγνώρισης για όλες τις δομές της γλώσσας καθώς και περιπτώσεις μη έγκυρων δομών τις οποίες ο ΣΑ θα πρέπει να αντιμετωπίζει με κάποιο μήνυμα συντακτικού σφάλματος
3. Ένα Makefile που θα επιτρέπει την αυτόματη μεταγλώττιση και εκτέλεση του κώδικά σας, όπως αυτό που παρέχεται στον πρότυπο κώδικα που κατεβάσατε
4. Το τεύχος τεκμηρίωσης της εργασίας σας σε μορφή Word ή PDF με όλα τα απαραίτητα περιεχόμενα (πχ. αναφορά αρμοδιοτήτων των μελών της ομάδας, ρητή αναφορά εάν ο κώδικάς σας κάνει compile και εάν εκτελείται σωστά, κλπ)

Αν θέλετε να σας δοθούν αναλυτικά σχόλια ανάδρασης για την προσπάθειά σας, φροντίστε να παρουσιαστείτε στο 3ο (προαιρετικό) εργαστήριο υποστήριξης ομάδων για την υλοποίηση του Μέρους Β' της εργασίας εξαμήνου.

ΜΕΡΟΣ Β-2: Σύνδεση κώδικα flex με κώδικα bison

Στο βήμα αυτό πρέπει:

1. Να συμπεριληφθούν στον κώδικα bison που ετοιμάσατε στο Μέρος Β-1 όσες εκφράσεις της γραμματικής δεν έχετε ήδη συμπεριλάβει
2. Να γίνει η κατάλληλη προσαρμογή του κώδικα bison ώστε ο συντακτικός σας αναλυτής να μπορεί να συνεργαστεί με ρουτίνα λεκτικής ανάλυσης που θα παραχθεί μέσω του flex
3. Να γίνει η κατάλληλη προσαρμογή του κώδικα flex που ετοιμάσατε στο Μέρος Α-3 ώστε αντί της δημιουργίας ανεξάρτητου λεκτικού αναλυτή να δημιουργείται ρουτίνα λεκτικής ανάλυσης για τον συντακτικό αναλυτή που θα προκύψει από την προσαρμογή του κώδικα flex
4. Ακολουθείστε τις οδηγίες του εργαστηρίου ώστε να δημιουργήσετε ένα Makefile που θα επιτρέπει την αυτόματη μεταγλώττιση και εκτέλεση του κώδικά σας για τη δημιουργία συντακτικού αναλυτή σε συνεργασία με ρουτίνα λεκτικής ανάλυσης
5. Αν μια δομή του αρχείου εισόδου δεν μπορεί να αναγνωριστεί ως έγκυρη έκφραση, τότε πρέπει να καλείται η ρουτίνα διαχείρισης λαθών SyntaxError για να αντιμετωπίσει το πρόβλημα με τη μέθοδο του πανικού και ο μετρητής σημαντικών λαθών (fatal errors) αυξάνεται κατά 1, ενώ η ανάλυση συνεχίζεται στην αρχή της επόμενης συμβολοσειράς εισόδου (εντολής)
6. Στο τέλος της συντακτικής ανάλυσης που πραγματοποιεί ο ΣΑ πρέπει να εμφανίζεται το αποτέλεσμα της ανάλυσης του πηγαίου κώδικα εισόδου (λέξεις που αναγνωρίστηκαν σωστά, εκφράσεις που αναγνωρίστηκαν σωστά και αν περιείχε λεκτικά ή/και συντακτικά σφάλματα και πόσα), ακολουθώντας αυστηρά τη σύνταξη:

ΣΩΣΤΕΣ ΛΕΞΕΙΣ:<πλήθος σωστών λέξεων>

ΣΩΣΤΕΣ ΕΚΦΡΑΣΕΙΣ:<πλήθος σωστών εκφράσεων>

ΛΑΘΟΣ ΛΕΞΕΙΣ:<πλήθος λανθασμένων λέξεων>

ΛΑΘΟΣ ΕΚΦΡΑΣΕΙΣ:<πλήθος λανθασμένων εκφράσεων>

Προσοχή: Δεν έχετε ακόμη διδαχθεί όλη την απαραίτητη ύλη για την μέθοδο πανικού και την καταμέτρηση των λάθους εκφράσεων. Χρησιμοποιώντας όμως το διαθέσιμο υλικό στο Open eClass για το εργαλείο bison μπορείτε να δοκιμάσετε να υλοποιήσετε τα ζητούμενα αυτά. Σε κάθε περίπτωση, εάν δεν τα καταφέρετε τελικά, θα έχετε τη δυνατότητα να ολοκληρώσετε την υλοποίησή σας αυτή στο μέρος Β-3 της εργασίας.

Ελέγξτε εξαντλητικά τον ΣΑ που θα δημιουργηθεί χρησιμοποιώντας ως είσοδο έναν πηγαίο κώδικα που θα περιέχει και σωστές αλλά και λανθασμένες λέξεις και εκφράσεις της γλώσσας. Η πληρότητα του παραπάνω ελέγχου αποτελεί βασικό στοιχείο στην αξιολόγηση της εργασίας.

Αναρτήστε έγκαιρα τον ολοκληρωμένο κώδικά σας, σωστά σχολιασμένο για να γίνεται κατανοητός, μαζί με τα αποτελέσματα των ελέγχων σας. Η κατάθεσή σας θα είναι ένα συμπιεσμένο αρχείο ZIP με όνομα που θα αρχίζει με τον κωδικό της ομάδας υποχρεωτικά και που θα περιλαμβάνει τα παρακάτω:

1. Τα αρχεία για τη δημιουργία του συντακτικού αναλυτή (κώδικες bison, flex, header)
2. Συνοδευτικά αρχεία εισόδου - εξόδου (input.txt - output.txt) που θα περιλαμβάνουν αρκετές περιπτώσεις αναγνώρισης για όλες τις δομές της γλώσσας καθώς και περιπτώσεις μη έγκυρων δομών τις οποίες ο ΣΑ θα πρέπει να αντιμετωπίζει με την μέθοδο πανικού (ή αν αυτό δεν καταστεί εφικτό, με κάποιο μήνυμα συντακτικού σφάλματος)
3. Ένα Makefile που θα επιτρέπει την αυτόματη μεταγλώττιση και εκτέλεση του κώδικά σας, όπως αυτό που παρέχεται στον πρότυπο κώδικα που κατεβάσατε
4. Το τεύχος τεκμηρίωσης της εργασίας σας σε μορφή Word ή PDF με όλα τα απαραίτητα περιεχόμενα (πχ. αναφορά αρμοδιοτήτων των μελών της ομάδας, ρητή αναφορά εάν ο κώδικάς σας κάνει compile και εάν εκτελείται σωστά, κλπ)

Αν θέλετε να σας δοθούν αναλυτικά σχόλια ανάδρασης για την προσπάθειά σας, φροντίστε να παρουσιαστείτε στο 4ο (προαιρετικό) εργαστήριο υποστήριξης ομάδων για την υλοποίηση του Μέρους Β' της εργασίας εξαμήνου.

ΜΕΡΟΣ Β-3: Διαχείριση λεκτικών και συντακτικών προειδοποιητικών λαθών

Για τη διαχείριση λεκτικών και συντακτικών προειδοποιητικών λαθών απαιτείται ο εμπλουτισμός των κανόνων της γραμματικής.

Για να πετύχετε το παραπάνω θα πρέπει:

- να ολοκληρώσετε την υλοποίησή σας για την καταμέτρηση των λανθασμένων εκφράσεων χρησιμοποιώντας πλέον την μέθοδο του πανικού στον ΣΑ σας
- να εμπλουτίσετε τους γραμματικούς κανόνες που έχετε δημιουργήσει στον κώδικα `bison` με κανόνες αναγνώρισης συνηθισμένων συντακτικών προειδοποιητικών λαθών ώστε να μπορεί γίνεται ανάνηψη από τα σφάλματα και η συντακτική ανάλυση να συνεχίζει την αναγνώριση στην αρχή της επόμενης συμβολοσειράς εισόδου (εντολής), εμφανίζοντας στο αρχείο εξόδου τα κατάλληλα μηνύματα λαθών και αυξάνοντας τους αντίστοιχους μετρητές λαθών.
- να εμπλουτίσετε τον κώδικα `flex` με επιπλέον κανόνες εντοπισμού προειδοποιητικών λαθών και αντιμετώπισής τους, με κατάλληλη δημιουργία αντίστοιχων κανονικών εκφράσεων, εμφανίζοντας στο αρχείο εξόδου τα κατάλληλα μηνύματα λαθών και αυξάνοντας τους αντίστοιχους μετρητές λαθών.

Παράδοση τελικής εργασίας

Παραδώστε τα παρακάτω σε συμπιεσμένη μορφή, με βάση συμπληρωματικές οδηγίες που θα σας δοθούν στο εργαστήριο:

1. το τεύχος της τεκμηρίωσης της εργασίας σας σε μορφή PDF με τα ακόλουθα:
 - ο εξώφυλλο (με ονοματεπώνυμο μελών ομάδας, αριθμός ομάδας και τμήμα), πίνακα περιεχομένων, εισαγωγικό κείμενο
 - τεκμηρίωση της γραμματικής (για λεκτική, συντακτική, σημασιολογική ανάλυση σε μορφή κανόνων EBNF) και του σχεδιασμού
 - τεκμηρίωση του κώδικα που δίνεται ως είσοδος στον `flex` και αυτού που δίνεται ως είσοδος στον `bison`
 - παρουσίαση ελέγχων με σχολιασμένα αποτελέσματα και τελικά σχόλια επί της εργασίας
 - αναφορά αρμοδιοτήτων όλων των μελών της ομάδας καθώς και αναφορά του αντίστοιχου υπεύθυνου εργασίας για κάθε ξεχωριστή ενότητα της εργασίας
 - ρητή αναφορά εάν ο κώδικάς σας κάνει `compile` με επιτυχία ή όχι, αν υπάρχουν προειδοποιητικά σφάλματα (`warnings`) και εάν εκτελείται ή όχι σωστά
2. τους (αναλυτικά σχολιασμένους) κώδικες αρχείων εισόδου για `flex` και `bison`
3. το αρχείο `Makefile` εκτέλεσης των παραπάνω
4. (προαιρετικά) ένα ξεχωριστό αρχείο Word ή PDF με τις αναλυτικές διορθώσεις όλων των προηγούμενων εργασιών
5. το αρχείο των εξαντλητικών δοκιμών σας με όνομα `input.txt` και το αντίστοιχο `output.txt`.
6. αν το κρίνετε απαραίτητο, κάποιο σημείωμα (απλό αρχείο κειμένου) με πιθανές παρατηρήσεις-προβλήματα που πρέπει οπωσδήποτε να προσέξει ο αξιολογητής της εργασίας

Καταληκτικές Ημερομηνίες παράδοσης

Τελική εργασία: ??/??/2021, 23:55 (Η ημερομηνία θα ανακοινωθεί στο μέλλον)

Οδηγίες για την κωδικοποίηση: Φροντίστε να ορίσετε κατάλληλη στρατηγική επίλυσης ώστε να μην εμπλακεί η εκτέλεση σε βρόχο ή εφαρμόστε δικές σας τακτικές για να τους αποφύγετε, προσθέστε προτεραιότητες μέσα στους κανόνες για την αποφυγή περιττών κινήσεων, εφαρμόσετε σωστά τις δεσμεύσεις των μεταβλητών ώστε η σύνταξη των κανόνων να είναι ταυτόχρονα ορθή αλλά και κομψή και τέλος προσθέστε σχόλια σε κάθε κανόνα που να περιγράφει το σκοπό του.

Οδηγίες για την τεκμηρίωση: Συντάξτε τεύχος τεκμηρίωσης με εξώφυλλο, περιεχόμενα, εισαγωγή, σώμα τεκμηρίωσης, δοκιμαστικές εκτελέσεις και παρουσίαση αποτελεσμάτων κατάλληλα σχολιασμένων ώστε να γίνονται κατανοητά.

Σημείωση:

Καταθέσεις εργασιών με όνομα συμπιεσμένου αρχείου που δεν αρχίζει με κωδικό ομάδας **ΔΕΝ ΘΑ ΓΙΝΟΝΤΑΙ ΔΕΚΤΕΣ**.

Διευκρινίζεται επίσης ότι όπου απαιτείται κώδικας ή αποτελέσματα:

1. Δεν θα γίνονται δεκτά *greeklish*, αλλά μόνο ελληνικά (ή μόνο σωστά αγγλικά).
2. Δεν θα γίνονται δεκτά *image snapshots / captures*, αλλά μόνο κείμενο κατάλληλα μορφοποιημένο (σχετικά με την αναδίπλωση γραμμών – *line wrapping*, ώστε να είναι εποπτικό και αναγνώσιμο).
3. Τα αποτελέσματα δοκιμών θα πρέπει να είναι σχολιασμένα ώστε να γίνονται κατανοητά. Σημειώνεται ότι οι δοκιμές πρέπει να είναι εξαντλητικές για να πιστοποιούν την ορθότητα και αρτιότητα του κώδικα.
4. Σε κάθε περίπτωση, θα πρέπει να δηλώσετε ρητά μέσα στο έγγραφο τεκμηρίωσης εάν ο κώδικάς σας κάνει ή όχι *compile* και εάν παράγει ή όχι σωστά/αποδεκτά αποτελέσματα. Επιπρόσθετα, τυχόν προβλήματα ή ελλείψεις κατά την εκτέλεση επιβάλλεται να αναφέρονται.

Για την καλύτερη κατανόηση των κριτηρίων αξιολόγησης της εργασίας σας, καλό θα ήταν να διαβάσετε προσεκτικά τη [ρουμπρίκα αξιολόγησης](#) που βρίσκεται στο eClass.