

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 **BNF/EBNF** και για τον τρόπο λειτουργίας των **συντακτικών και σημασιολογικών αναλυτών** που έχετε 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 αυτή.

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 BNF/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 BNF/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.

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

ΜΕΡΟΣ B-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](#).

Αφού κατανοήσετε τη λειτουργία της γεννήτριας φροντίστε να διορθώσετε και να ολοκληρώσετε τον συντακτικό αναλυτή αντικαθιστώντας τα **"FILL ME"** με τον σωστό κώδικα bison προκειμένου να επιτύχετε την αναγνώριση όσων περισσότερων εκφράσεων της γλώσσας του εργαστηρίου μπορείτε.

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 θα πρέπει να εμφανίζεται στο αρχείο εξόδου αριθμημένη.

Προσοχή: Το βήμα αυτό σχεδιάστηκε για δική σας εξάσκηση **χωρίς να χρειάζεται να καταθέσετε κάτι σχετικό στο eClass ή να συμπεριλάβετε το βήμα αυτό στο τελικό σας παραδοτέο!**

ΜΕΡΟΣ B-2: Σύνδεση κώδικα flex με κώδικα bison

Στο βήμα αυτό πρέπει:

1. Να συμπεριληφθούν στον κώδικα bison που ετοιμάσατε στο Μέρος B-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 A-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. Να ακολουθήσετε 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. 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 καταμέτρηση των λάθος εκφράσεων. Για το μέρος B-2 μπορείτε προαιρετικά, χρησιμοποιώντας το διαθέσιμο υλικό στο Open eClass για το εργαλείο bison, να δοκιμάσετε να υλοποιήσετε το ζητούμενο αυτό χρησιμοποιώντας τη μέθοδο πανικού. Σε κάθε περίπτωση, η καταμέτρηση των λάθος εκφράσεων απαιτείται μονάχα για το μέρος B-3 της εργασίας που άλλωστε αποτελεί και το μόνο παραδοτέο για το B' μέρος της εργασίας εξαμήνου.

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

Προσοχή: Το βήμα αυτό σχεδιάστηκε για να προετοιμάσετε/υλοποιήσετε κατάλληλα τον ΣΑ σας (που ξεκινήσατε από το μέρος B-1), συνδέοντάς τον με τον ΛΑ του μέρους A-3. Μολονότι **ΔΕΝ χρειάζεται να καταθέσετε κάτι στην παρούσα φάση, όλο το υλικό που θα προετοιμάσετε για το μέρος B-2 θα χρειαστεί να κατατεθεί μαζί με τα επιπρόσθετα ζητούμενα του μέρους B-3 στο επόμενο εργαστηριακό μάθημα!**

ΜΕΡΟΣ B-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 οπωσδήποτε την μέθοδο του πανικού στον ΣΑ σας. 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`) να αυξάνεται κατά 1, ενώ η ανάλυση να συνεχίζεται στην αρχή της επόμενης συμβολοσειράς εισόδου (εντολής).
- 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 (παραδοτέο μέρους B-3)

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 Word, Writer ή PDF με τα ακόλουθα:
 - εξώφυλλο (με ονοματεπώνυμο μελών ομάδας, αριθμό ομάδας και τμήμα), πίνακα περιεχομένων, εισαγωγικό κείμενο
 - τεκμηρίωση της γραμματικής (για λεκτική, συντακτική, σημασιολογική ανάλυση) και του σχεδιασμού
 - τεκμηρίωση του κώδικα που δίνεται ως είσοδος στον `flex` και αυτού που δίνεται ως είσοδος στον `bison`
 - παρουσίαση εξαντλητικών ελέγχων με σχολιασμένα αποτελέσματα και τελικά σχόλια επί της εργασίας
 - αναλυτική αναφορά αρμοδιοτήτων όλων των μελών της ομάδας καθώς και ονομαστική αναφορά στα μέλη που αποχώρησαν από την ομάδα σας (αν υπάρχουν) και σε ποιο μέρος της εργασίας συνέβη αυτό
 - ρητή αναφορά εάν ο κώδικάς σας κάνει `compile` ("περνάει" από τα εργαλεία `bison`, `flex` και `gcc`) με επιτυχία ή όχι, αν υπάρχουν προειδοποιητικά σφάλματα (`warnings`) και εάν τελικά εκτελείται ή όχι σωστά
2. τους (αναλυτικά σχολιασμένους) κώδικες για `flex` (αρχείο `.l`) και `bison` (αρχείο `.y`)

3. το αρχείο Makefile που θα επιτρέπει την αυτόματη μεταγλώττιση και εκτέλεση του κώδικά σας, όπως αυτό που παρέχεται στον πρότυπο κώδικα που κατεβάσατε
4. (προαιρετικά) ένα ξεχωριστό αρχείο Word, Writer ή PDF με τις αναλυτικές διορθώσεις (**και μόνο αυτές**) των προηγούμενων εργασιών (A-2 και A-3). Προσοχή: Πλήρης επανυποβολή των προηγούμενων παραδοτέων δεν θα αξιολογηθεί!
5. το αρχείο (ή τα αρχεία) των εξαντλητικών δοκιμών σας και το αντίστοιχο αρχείο (ή αρχεία) εξόδου. Υπενθυμίζεται ότι τα συνοδευτικά αρχεία εισόδου - εξόδου (input.txt - output.txt) θα πρέπει να 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)
6. (προαιρετικά) αν το κρίνετε απαραίτητο, κάποιο σημείωμα (απλό αρχείο κειμένου) με πιθανές παρατηρήσεις-προβλήματα που πρέπει οπωσδήποτε να προσέξει ο αξιολογητής της εργασίας πριν ή κατά την εκτέλεση του κώδικά σας

Σημείωση: Μπορείτε προαιρετικά -- κι αφού πρώτα ολοκληρώσετε τα βασικά ζητούμενα της εργασίας -- να προχωρήσετε στην υλοποίηση και τεκμηρίωση των παρακάτω: πίνακας συμβόλων (symbol table), αφηρημένο συντακτικό δέντρο (abstract syntax tree), ενδιάμεσος κώδικας.

Οδηγίες για την τεκμηρίωση: Συντάξτε τεύχος τεκμηρίωσης με εξώφυλλο, περιεχόμενα, εισαγωγή, σώμα τεκμηρίωσης, εξαντλητικές 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.