

Αρχές Γλωσσών Προγραμματισμού & Μεταφραστών**Τμήμα Μηχανικών Η/Υ & Πληροφορικής****Πανεπιστήμιο Πατρών****Εαρινό Εξάμηνο 2021****Διδάσκοντες: Ι. Γαροφαλάκης, Σ. Σιούτας****Εργαστηριακή Άσκηση**

Σκοπός της εργαστηριακής άσκησης είναι να εξοικειωθείτε με την περιγραφή μιας γλώσσας σε μορφή BNF, βασικές έννοιες των μεταγλωττιστών και τελικά να υλοποιήσετε δύο από τα τμήματα ενός μεταγλωττιστή, έναν συντακτικό και ένα λεκτικό αναλυτή με τη χρήση των εργαλείων Bison και Flex. Η γλώσσα που θα αναλυθεί είναι μια ψευδογλώσσα, η οποία ακολουθεί τη λογική της γλώσσας C, και περιγράφεται αμέσως μετά.

- ✓ Κάθε πρόγραμμα **ξεκινά** με τη δεσμευμένη λέξη **PROGRAM** η οποία ακολουθείται από μια ακολουθία χαρακτήρων που υποδηλώνουν το όνομα του προγράμματος. Μετά το όνομα του προγράμματος πρέπει υποχρεωτικά να υπάρχει **αλλαγή γραμμής**.

- ✓ Στη συνέχεια, **υπάρχει (προαιρετικός) ορισμός συναρτήσεων**. Οι συναρτήσεις ορίζονται ως εξής:

- ο **Ξεκινούν** με τη δεσμευμένη λέξη **FUNCTION** η οποία ακολουθείται από το **όνομά της** συνάρτησης και στη συνέχεια, **εντός παρενθέσεων**, από **λίστα των παραμέτρων** χωρισμένων με κόμμα.

VARS

```
CHAR var1, var2;  
CHAR var3, var4;  
INTEGER var5;
```

- ο Έπειτα, μετά από αλλαγή γραμμής, υπάρχει το σώμα της συνάρτησης. Στο σώμα της συνάρτησης αρχικά έχουμε την **προαιρετική δήλωση μεταβλητών**. Η δήλωση γίνεται με τη δεσμευμένη λέξη **VARS**. Κατόπιν, πρώτα ορίζεται ο **τύπος δεδομένων των μεταβλητών (CHAR και INTEGER)** ακολουθούμενος από λίστα με τα ονόματα των μεταβλητών (περισσότερες από μια μεταβλητές χωρίζονται με ","). Στο τέλος πρέπει να υπάρχει ";".

- ο Πρέπει να υπάρχει η δυνατότητα ορισμού **πινάκων** με τη δήλωση του μήκους (υποχρεωτικά θετικός ακέραιος) εντός **[]**.

- Π.χ. CHAR var1,var2,var3[10],var4;

- ο Στη συνέχεια, ακολουθούν οι **εντολές** του προγράμματος, που αναλύονται στη συνέχεια.

- ο Το τέλος της συνάρτησης ορίζεται ρητά με τη δεσμευμένη λέξη **END_FUNCTION**. Πριν κλείσει η συνάρτηση, θα πρέπει υποχρεωτικά να επιστρέφεται μια τιμή με τη δεσμευμένη λέξη **RETURN** ακολουθούμενη από την τιμή, η οποία μπορεί να είναι είτε όνομα μεταβλητής, είτε κυριολεκτικό (literal).

- ✓ Αμέσως μετά τον ορισμό των συναρτήσεων, ακολουθεί το **κύριο μέρος του προγράμματος** το οποίο περικλείεται από τις **δεσμευμένες λέξεις STARTMAIN και ENDMAN**. Στο κύριο μέρος του προγράμματος, αρχικά έχουμε την **προαιρετική δήλωση μεταβλητών** (ακριβώς όπως περιγράφηκε παραπάνω για τις συναρτήσεις), ενώ στη συνέχεια μπορούν να εμφανίζονται **με οποιαδήποτε σειρά οποιοσδήποτε από τις εντολές του προγράμματος** που περιγράφονται παρακάτω.

- ✓ Οι **Εντολές Προγράμματος** μπορούν να εμφανίζονται με οποιαδήποτε σειρά και διακρίνονται σε:

◆ *Εντολή ανάθεσης*

Έχουν τη μορφή **<μεταβλητή>=<έκφραση>**;, όπου η έκφραση μπορεί να είναι ένα κυριολεκτικό ή μια σύνθετη παράσταση από συνδυασμό κυριολεκτικών, μεταβλητών και παρενθέσεων ή κλήση συνάρτησης. Επιπλέον, η σύνθετη παράσταση μπορεί να είναι οποιαδήποτε αριθμητική παράσταση που περιλαμβάνει τις πράξεις **+**, **-**, **^**, *****, **/**.

Π.χ.:

var1 = 2;

var2 = var3 + 5 * 2^(3-7/(var1+var3));

var4 = var3;

var5 = function1(var1,var2);

Οι μεταβλητές μπορούν να εμφανίζονται σε αυτό το τμήμα του προγράμματος ανεξάρτητα εάν έχουν δηλωθεί στο προηγούμενο τμήμα δήλωσης μεταβλητών ή όχι. Ομοίως και για τις συναρτήσεις, δεν είναι υποχρεωτικό να έχει προηγηθεί δήλωσή τους.

◆ *Εντολές βρόχου*

Υπάρχουν 2 είδη εντολών βρόχου. Το πρώτο είδος έχει την ακόλουθη μορφή:

WHILE (συνθήκη)

... εντολές προγράμματος ...

ENDWHILE

Όπου **συνθήκη** είναι οποιαδήποτε λογική έκφραση που περιέχει τους συγκριτικούς τελεστές **>**, **<**, **==**, **!=** και τους λογικούς **AND** και **OR**.

Το δεύτερο είδος έχει την ακόλουθη μορφή:

FOR counter:=1 TO 100 STEP 2 (τα νούμερα είναι ενδεικτικά)

... εντολές προγράμματος ...

ENDFOR

◆ *Εντολές ελέγχου*

Υπάρχουν 2 είδη εντολών ελέγχου. Το πρώτο είδος έχει την ακόλουθη μορφή:

IF (συνθήκη) THEN

... εντολές προγράμματος ...

ELSEIF

... εντολές προγράμματος ...

ELSEIF

... εντολές προγράμματος ...

...

ELSE

... εντολές προγράμματος ...

ENDIF

Η εμφάνιση των ELSEIF και ELSE είναι προαιρετική. Όπου **συνθήκη** είναι οποιαδήποτε λογική έκφραση που περιέχει τους συγκριτικούς τελεστές >, <, ==, != και τους λογικούς **AND** και **OR**.

Το δεύτερο είδος έχει την ακόλουθη μορφή:

SWITCH(<έκφραση>)**CASE(<έκφραση>):**

... εντολές προγράμματος...

CASE(<έκφραση>):

... εντολές προγράμματος...

...

DEFAULT:

... εντολές προγράμματος...

ENDSWITCH

Η επιλογή DEFAULT είναι προαιρετική

◆ *Εντολή εκτύπωσης στην οθόνη*

Έχει τη μορφή

PRINT("κείμενο "[,var1]);

(Δηλαδή κείμενο εντός "" κι αμέσως μετά προαιρετικά όνομα μεταβλητής)

◆ *Εντολή τερματισμού βρόχου*

Έχει τη μορφή

BREAK;

◆ *Σχόλια μιας γραμμής*

Είναι οτιδήποτε υπάρχει μετά τον χαρακτήρα % και μέχρι το τέλος της ίδιας γραμμής.

- ✓ Εάν δεν αναφέρεται ρητά κάτι διαφορετικό, τα κενά και οι χαρακτήρες αλλαγής γραμμής δεν αποτελούν μέρος της γλώσσας και δεν παίζουν κανένα ρόλο στη σύνταξη.
- ✓ Όλες οι δεσμευμένες λέξεις της γλώσσας είναι υποχρεωτικά με κεφαλαίους χαρακτήρες.
- ✓ Τα ονόματα των μεταβλητών και των συναρτήσεων ακολουθούν τους κανόνες ονοματοδοσίας της C.

Ερωτήματα

1. (60%)

- a. Δώστε σε BNF τον συντακτικό ορισμό της ψευδογλώσσας.
- b. Χρησιμοποιώντας τα προγράμματα Flex και Bison, υλοποιήστε έναν λεξικό και συντακτικό αναλυτή, ο οποίος θα παίρνει ως είσοδο ένα αρχείο γραμμένο στη ψευδογλώσσα που περιγράφηκε παραπάνω και θα ελέγχει αν το πρόγραμμα είναι συντακτικά ορθό. Το πρόγραμμά σας θα καλείται από τη γραμμή εντολών ως εξής:

prompt> myParser.exe file.c

και θα επιστρέφει το ίδιο το πρόγραμμα και ένα διαγνωστικό μήνυμα για το αν ήταν ορθώς γραμμένο, ή κατάλληλο μήνυμα σφάλματος (πρέπει να φαίνεται η γραμμή όπου υπάρχει το σφάλμα).

2. (10%)

Επεκτείνετε τον παραπάνω αναλυτή προσθέτοντας τα παρακάτω:

Δήλωση τύπου δεδομένων χρήστη / Δήλωση δομής:

Η δήλωση του τύπου δεδομένων χρήση / δήλωση δομής γίνεται πριν τον (προαιρετικό) ορισμό των συναρτήσεων.

Οι τύποι δεδομένων χρήστη ορίζονται με τον ακόλουθο τρόπο:

- Ξεκινούν με τη δεσμευμένη λέξη **STRUCT** η οποία ακολουθείται από το όνομά του τύπου.
- Έπειτα, μετά από αλλαγή γραμμής, υπάρχουν οι δηλώσεις μεταβλητών. Η δήλωση γίνεται με τη δεσμευμένη λέξη **VARS**. Κατόπιν, πρώτα ορίζεται ο τύπος δεδομένων των μεταβλητών (**CHAR** και **INTEGER**) ακολουθούμενος από το όνομα της μεταβλητής. Πρέπει να υπάρχει η δυνατότητα ορισμού πινάκων με τη δήλωση του μήκους (υποχρεωτικά θετικός ακέραιος) εντός []. Στο τέλος πρέπει να υπάρχει ";".
- Το τέλος του ορισμού του τύπου δεδομένων ορίζεται ρητά με τη δεσμευμένη λέξη **ENDSTRUCT**.
- Προαιρετικά, πριν από τη δεσμευμένη λέξη **STRUCT** μπορεί να εμφανίζεται η δεσμευμένη λέξη **TYPDEF** (π.χ. **TYPDEF STRUCT <όνομα τύπου>**). Στη συνέχεια, ακολουθούν οι δηλώσεις μεταβλητών. Το τέλος του ορισμού σε αυτή την περίπτωση ορίζεται ως **<όνομα τύπου> ENDSTRUCT**.

Αν υλοποιηθεί το παρόν ερώτημα, θα πρέπει να γίνει κατάλληλη μετατροπή στον αναλυτή και στο κομμάτι που αφορά τη δήλωση μεταβλητών, ώστε να αναγνωρίζονται οι τύποι δεδομένων που θα δηλωθούν από τη χρήση από τη γραμματική σας.

Ιδανικά πρέπει να αναγνωρίζεται οποιοδήποτε όνομα τύπου, ωστόσο για ευκολία μπορείτε να υλοποιήσετε μόνο έναν τύπο, παραδοχή την οποία πρέπει να αναλύετε στην αναφορά σας.

3. (20%)

Κάντε τις απαραίτητες αλλαγές στα παραπάνω, ώστε να γίνεται έλεγχος σωστής δήλωσης των μεταβλητών που χρησιμοποιούνται οπουδήποτε στο πρόγραμμα. Επιπλέον, να ελέγχεται ότι όπου καλούνται συναρτήσεις αυτές έχουν οριστεί προηγουμένως. Διαφορετικά η διαδικασία της ανάλυσης να διακόπτεται με εμφάνιση

1
2

σχετικού σφάλματος.

4. (10%)

Κάντε τις απαραίτητες αλλαγές στο προηγούμενο ερώτημα, ώστε να υποστηρίζονται τα **σχόλια πολλαπλών γραμμών που έχουν την ακόλουθη μορφή:**

/* σχόλια

ή

/* σχόλια */

....

σχόλια */

Σημείωση: Να μην γίνει χρήση κανονικής έκφρασης για την επίλυση του συγκεκριμένου ερωτήματος.

Παραδοτέα

- ❖ **Γραπτή Αναφορά** που περιλαμβάνει
 - Τα στοιχεία των μελών της ομάδας (Ονοματεπώνυμο, ΑΜ, Έτος, e-mail)
 - Την περιγραφή της γραμματικής της γλώσσας σε BNF.
 - Τα **ΤΕΛΙΚΑ** αρχεία περιγραφής της γλώσσας, τα οποία δίνονται ως **είσοδος** στα Flex και Bison
 - Screenshots παραδειγμάτων εφαρμογής (τουλάχιστον ένα παράδειγμα για κάθε ερώτημα, για επιτυχημένη και ανεπιτυχή ανάλυση)
 - Σχόλια - Παραδοχές που έγιναν για την ανάπτυξη της εργασίας
- ❖ Συμπίεσμένα σε **ένα αρχείο (zip)** τα:
 - Την παραπάνω γραπτή αναφορά.
 - Τα **ΤΕΛΙΚΑ** αρχεία εισόδου για το flex και το bison.
 - Τον **ΤΕΛΙΚΟ** κώδικα σε C (και το .h αρχείο) που παράγεται από τα δύο προγράμματα και ο εκτελέσιμος κώδικας του parser.
 - Τα αρχεία που δόθηκαν σαν είσοδο στον parser για να ελεγχθεί η σωστή λειτουργία του.

Το αρχείο zip πρέπει να έχει όνομα τους αριθμούς μητρώου των ατόμων της ομάδας διαχωρισμένους με _ και από τον μικρότερο στο μεγαλύτερο (π.χ. 1000_1543_2788_3972.zip), και να υποβληθεί (ΥΠΟΧΡΕΩΤΙΚΑ) στο eclass του μαθήματος και στο μήνυμα να αναφέρονται τα ονοματεπώνυμα, το έτος και οι αντίστοιχοι αριθμοί μητρώου των μελών της ομάδας, καθώς και οι e-mail διευθύνσεις όλων των μελών της ομάδας

Διευκρινίσεις

- ❖ Για τη λειτουργία των εργαλείων Flex και Bison μπορείτε να βρείτε πληροφορίες στο eclass του μαθήματος.
- ❖ Η άσκηση θα γίνει σε ομάδες έως 4 ατόμων.
- ❖ Η συμβολή της στον τελικό βαθμό είναι 30%.
- ❖ Η άσκηση είναι υποχρεωτική και πρέπει ο βαθμός της να είναι τουλάχιστον 5.
- ❖ Τελικές ημερομηνίες παράδοσης είναι οι ημερομηνίες γραπτών εξετάσεων περιόδου Ιουνίου και Σεπτεμβρίου αντίστοιχα. Αναλόγως θα καθοριστεί και η ημερομηνία της προφορικής εξέτασης της εργασίας.
- ❖ Τυχόν απορίες ή υποδείξεις μπορείτε να τις αναρτήσετε στο χώρο “Συζητήσεις” στην σελίδα του μαθήματος στο eclass