*Θέμα 5*

Η παρουσίαση αυτή έχει ως σκοπό την αναλυτική εξήγηση του 5ου θέματος της εργασίας στο μάθημα των μεταγλωττιστών , καθώς και την ανάλυση στον τρόπο σκέψης και πως αυτό μεταφράζεται σε κώδικα.

Το "Thema5" αρχείο αντιπροσωπεύει τον κώδικα που έχει γραφτεί χρησιμοποιώντας visual studio code και το "a" αρχείο το εκτελέσιμο κομμάτι του παραπάνω κώδικα. Tο πρόγραμμα έχει γραφτεί σε γλώσσα φλεξ με βάση αυτά που διδαχτήκαμε στο φετινό εργαστήριο του εξαμήνου μας.

Ο σκοπός του προγράμματος μας είναι να **δέχεται μία φράση που θα αντιπροσωπεύει ένα γεωμετρικό σχήμα και δίπλα τα σημεία** (ή αλλιώς τις κορυφές) που το συνοδεύουν και να κρίνει αν αυτή η έκφραση είναι γεωμετρικά αποδεκτή.

Να σημειωθεί εδώ ότι τα γράμματα που μπορούν να δοθούν ως κορυφές είναι: **A,B,C,D,E,F,G,H**

Για να ξεκινήσουμε είναι σημαντικό να κατανοήσουμε μερικά πράγματα πρώτα:

Για παράδειγμα **ένα τρίγωνο συνοδεύεται από 3 μοναδικές κορυφές** (ή σημεία) ,και μόνο 3,αλλιώς δεν λέγεται τρίγωνο. Οπότε μια φράση του τύπου ***«triangle ABC»*** είναι γεωμετρικά ***σωστή!***

Μια φράση του τύπου ***«triangle ABCD»*** είναι σαφώς ***λανθασμένη* καθώς ένα τρίγωνο δεν μπορεί να έχει πέρα από 3 κορυφές.**

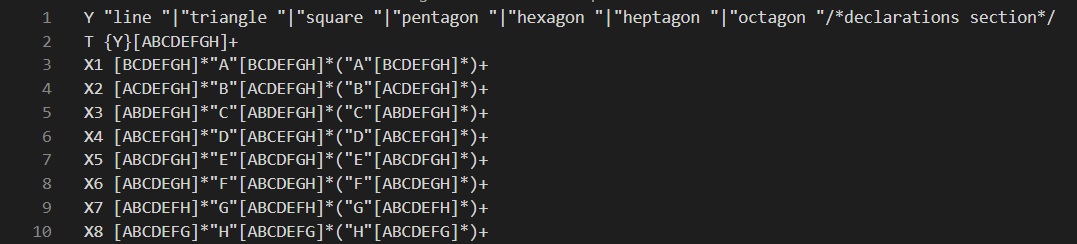
Μια ακόμη ***λανθασμένη*** έκφραση θα ήταν ***«triangle AAC»*** καθώς **η κάθε κορυφή πρέπει να είναι μοναδική.**

**Με βάση τα παραπάνω καταλήγουμε σε δύο πολύ απλά συμπεράσματα:**

1. **Κάθε ν-γώνο πρέπει να συνοδεύεται από τον αντίστοιχο αριθμό κορυφών του, δηλαδή ν.**
2. **Κάθε κορυφή (δηλαδή γράμμα) μπορεί να εμφανιστεί μόνο ΜΙΑ ΦΟΡΑ.**

Οπότε το πρόγραμμα μας θα πρέπει να τα απορρίπτει αυτά!

**Τμήμα Δηλώσεων**



Στη μεταβλητή Y εκχωρούμε όλες τις αποδεχτές φράσεις **σχημάτων** που μπορούν **να προκύψουν με 2 έως 8 γράμματα!**

**\*λείπει η λέξη «point » ,αυτό είναι επειδή το Y θα χρησιμοποιηθεί αργότερα ως ελεγχτείς για πιθανόν διπλά γράμματα (διπλά ή περισσότερα) ή αλλιώς Duplicates και σε αυτό θα ήταν περιττός ο έλεγχος καθώς έχει max 1 γράμμα\***

Στη μεταβλητή T αποθηκεύουμε την έκφραση του Y με δίπλα όλα τα πιθανά γράμματα συνοδευόμενα από ένα **+.**

Αυτό σημαίνει «μία από τις εκφράσεις του Y,μαζί με ένα ή περισσότερα απ’ τα γράμματα ABCDEFGH»

\*Αυτό θα αξιοποιηθεί μετά για να αναγνωριστεί αν ο χρήστης δίνει άγνωστους για το πρόγραμμα χαρακτήρες\*

Οι μεταβλητές X1,X2,…,X8 έχουν παρόμοια σημασία, ο X1 δέχεται την φράση:

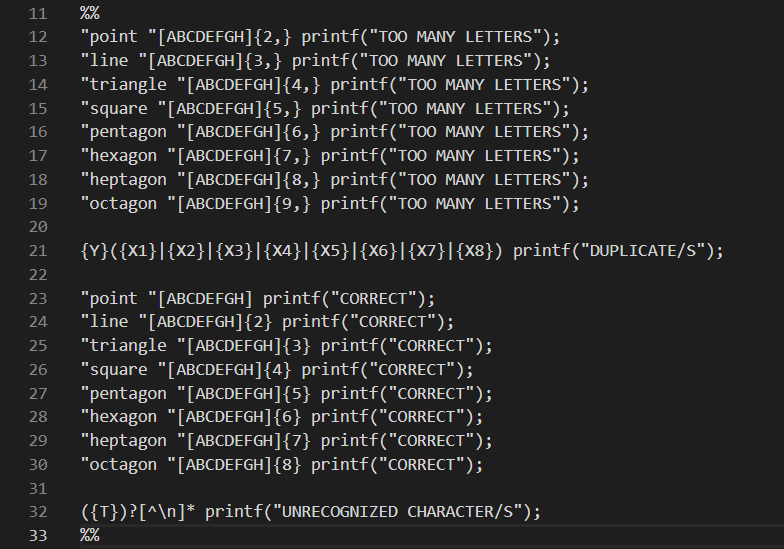
≪οτιδήποτε από τα BCDEFGH μηδέν έως άπειρες φορές, A, οτιδήποτε από τα BCDEFGH μηδέν έως άπειρες φορές≫. Μετά από αυτό, το: ≪Α, οτιδήποτε από τα BCDEFGH μηδέν έως άπειρες φορές≫ μπορεί να επαναληφθεί από ΜΙΑ έως άπειρες φορές.

Αυτό με άλλα λόγια σημαίνει δύο ή περισσότερα Άλφα, σε μία έκφραση που περιέχει A-H.

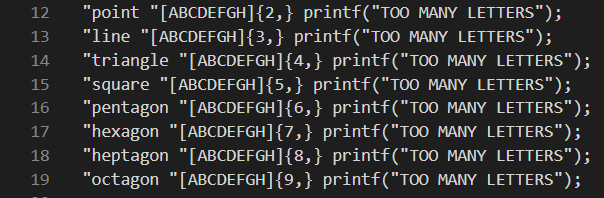
Η μεταβλητή Χ1 θα αξιοποιηθεί για να βρούμε αν υπάρχουν τυχόν παραπάνω από 1 Άλφα.

Οι μεταβλητές X2,X3,…,X8 κάνουν την ίδια ακριβώς διεργασία, απλά η κάθε μία για το κάθε αντίστοιχο γράμμα B,C,…,H και θα αξιοποιηθούν με τον αντίστοιχο τρόπο.

**ΤΜΗΜΑ ΕΛΕΓΧΟΥ ΚΑΙ ΕΚΤΕΛΕΣΗΣ**



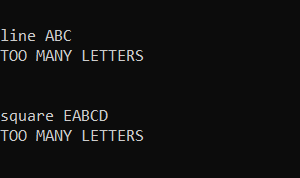
Την παραπάνω φώτο θα την χωρίσουμε σε 4 μέρη :



Αυτό είναι το τμήμα του προγράμματος που θα ελέγξει αν υπάρχουν τυχόν «περισσευούμενα» γράμματα στην έκφραση.

«**Κάθε ν-γώνο πρέπει να συνοδεύεται από τον αντίστοιχο αριθμό κορυφών του, δηλαδή ν.**

Η γραμμή 12 ελέγχει αν το μήνυμα του χρήστη ξεκινάει με «point +οποιοδήποτε από τα επιτρεπτά γράμματα **δύο ή περισσότερες φορές**» ,αν αυτό ισχύει θα εμφανίσει το μήνυμα «TOO MANY LETTERS» στο χρήστη, ενημερώνοντας τον πως έχει χρησιμοποιήσει **περισσότερα** απ’ τα γράμματα που αντιστοιχούν στη λέξη.



Το ίδιο ακριβώς συμβαίνει και στις υπόλοιπες περιπτώσεις, αντίστοιχα με τη λέξη βέβαια. (3 ή περισσότερα για το line,4 ή περισσότερα για το triangle,..)



Αυτή ίσως είναι και η πιο περίπλοκη γραμμή του προγράμματος, καθώς όλο της το νόημα είναι «μαζεμένο»..

**Αφού κάναμε τον έλεγχο για «περισσευούμενα» γράμματα, τώρα θα ελέγξουμε για διπλά,τριπλά,κλπ!**

Ο κανόνας {Y} αντιστοιχεί σε όλες τις αποδεκτές λέξεις/σχήματα:

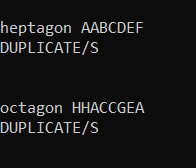


(εκτός από το **«point »**)

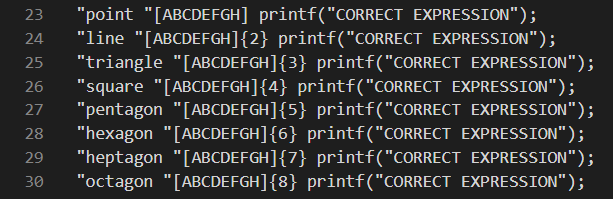
Ο κάθε κανόνας {X1},{X2},… αντιστοιχεί στον **έλεγχο κάθε γράμματος για το αν υπάρχει παραπάνω από μία φορά στην έκφραση** και τα | ανάμεσα αντιπροσωπεύουν τα «ή» μεταξύ των ελέγχων!

Η γενική σημασία της πρότασης αυτής είναι:

«μία από τις λέξεις του Υ + (το 'Α' δύο ή παραπάνω φορές Ή το 'Β' δύο ή παραπάνω φορές Ή … Ή το 'Η' δύο ή παραπάνω φορές» που σημαίνει **αν εμφανιστεί οποιοδήποτε γράμμα δύο ή περισσότερες φορές** θα εμφανίσει το μήνυμα «DUPLICATE/S».



\*Στην αρχή**, ελέγχεται αν έχουμε τον σωστό αριθμό γραμμάτων** για την κάθε λέξη. Σε αυτή τη περίπτωση έχουμε, αλλά υπάρχουν duplicates.\*



Αφού έχουν ολοκληρωθεί όλοι οι παραπάνω έλεγχοι, έφτασε η ώρα για τα αποτελέσματα μας..

Εφόσον τώρα ξέρουμε ότι ΔΕΝ υπάρχουν διπλά γράμματα ΚΑΙ για κάθε λέξη έχουμε τον αντίστοιχο αριθμό γραμμάτων μπορούμε να κάνουμε τον εξής έλεγχο!

Η γραμμή 23 σημαίνει:

«point + οποιοδήποτε από τα επιτρεπόμενα γράμματα»

Και η γραμμή 24 αντίστοιχα σημαίνει:

«line + οποιοδήποτε δύο από τα επιτρεπόμενα γράμματα»

Και η κάθε γραμμή βέβαια αντιστοιχεί στο δικό της αριθμό γραμμάτων:

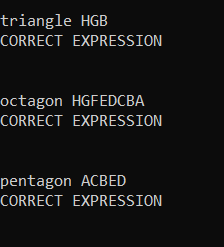
25: «triangle + οποιοδήποτε τρία από τα επιτρεπόμενα τρία γράμματα»

26: «square + οποιοδήποτε τέσσερα από τα επιτρεπόμενα γράμματα»

. . .

30: «octagon+ όλα τα επιτρεπόμενα γράμματα»

Μόλις βρεθεί η περίπτωση που αντιστοιχεί στο μήνυμα του χρήστη θα εμφανίσει το μήνυμα «CORRECT EXPRESSION» εάν η έκφραση που έδωσε ήταν σωστή.





Η έκφραση στη γραμμή 22 με απλά λόγια θα μπορούσε να εκφραστεί *οτιδήποτε άλλο(εκτός από τις αποδεκτές εκφράσεις βέβαια)*

Το {Τ} αντιστοιχεί στην εξής έκφραση:  που σημαίνει:

«μία απ’ τις εκφράσεις του Υ(line,triangle,...) + τα γράμματα ABCDEFGH μία ή περισσότερες φορές»

Το ({Τ})? σημαίνει:

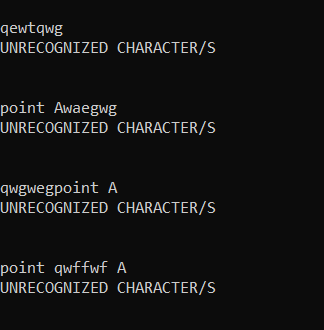
«(μία απ’ τις εκφράσεις του Υ(line,triangle,...) + κάποια από τα γράμματα ABCDEFGH μία ή περισσότερες φορές) από μηδέν μέχρι μία φορές»

Το [^\n]\* σημαίνει:

«(οτιδήποτε άλλο εκτός από αλλαγή γραμμής) από μηδέν έως άπειρες φορές»

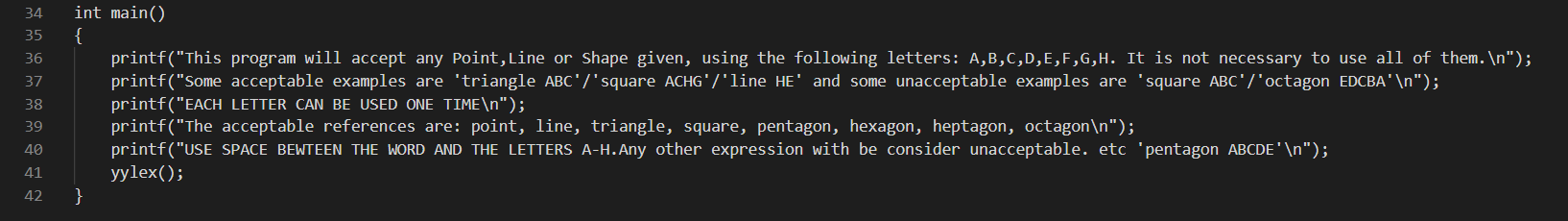
Οπότε πιο απλοϊκά η τελική έκφραση ({Τ})? [^\n]\* σημαίνει:

«μία ή καμία (πιθανών) δεκτή πρόταση + οτιδήποτε άλλο από μηδέν έως άπειρες φορές εκτός από αλλαγή γραμμής»



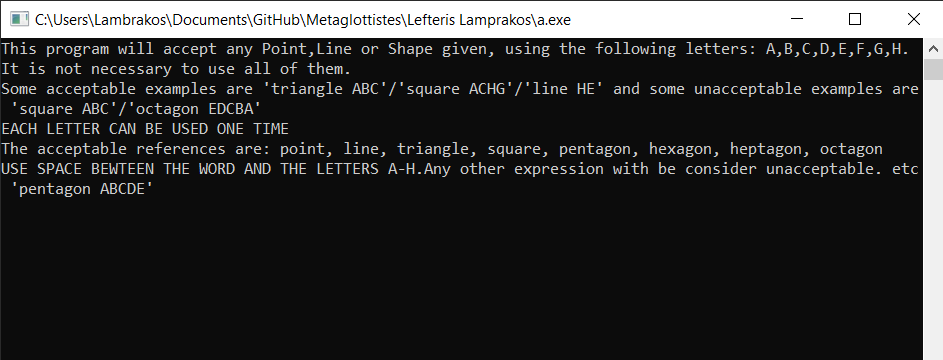
Όπως βλέπουμε οποιαδήποτε άλλη δήλωση πέρα από τις αποδεκτές θα απορριφθεί εμφανίζοντας το μήνυμα «UNRECOGNIZED CHARACTER/S», ακόμα και αν εμπεριέχει αποδεκτή έκφραση σε κάποιο σημείο της.

**ΕΜΦΑΝΙΣΕΙ ΜΗΝΥΜΑΤΩΝ + ΜΑΙΝ**



Το τμήμα της int main() εμπεριέχει όλα τα μηνύματα που θα εμφανιστούν στο χρήστη όταν ξεκινήσει το εκτελέσιμο πρόγραμμα καθώς και την εντολή yylex(); η οποία είναι υπεύθυνη για τη αποδοχή έκφρασης από τον χρήστη. Η εκτέλεση των μηνυμάτων γίνεται με τη χρήση της εντολής printf("μήνυμα");

Όταν ο χρήστης εκτελέσει το πρόγραμμα ,το εκτελέσιμο αρχείο θα έχει τη μορφή:



**Αξίζει να σημειωθεί εδώ ότι η φλεξ είναι περιορισμένη γλώσσα και όταν χρησιμοποιούμε τον όρο «οτιδήποτε» εννοούμε τους χαρακτήρες ASCII οι οποίοι είναι:**

* Τους αριθμούς 0-9
* Τα γράμματα του αγγλικού αλφάβητου (πεζά και κεφαλαία) A-Z και a-z
* Οι χαρακτήρες !, #, $, %, &, ΄,( ,) , { ,} ,[ ,],…

Ο πίνακας παρουσιάζεται παρακάτω

