

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Εργασία Μεταγλωττιστές

3 Εξαμηνο – Ακαδιμαικο ετος 2017-2018

Π16036 - ΙΩΑΝΝΙΔΗΣ ΠΑΝΑΓΙΩΤΗΣ

Π16112 - ΠΑΡΑΒΑΝΤΗΣ ΑΘΑΝΑΣΙΟΣ

2018

Περιεχόμενα

[2 Άσκηση Α 2](#_Toc504330740)

[2.1 Εισαγωγή 2](#_Toc504330741)

[2.2 Επίλυση 2](#_Toc504330742)

[2.3 Εκτέλεση 3](#_Toc504330743)

[3 Άσκηση Β 6](#_Toc504330744)

[3.1 Εισαγωγή 6](#_Toc504330745)

[3.2 Επίλυση 6](#_Toc504330746)

[3.3 Εκτέλεση 7](#_Toc504330747)

[4 Άσκηση Γ 9](#_Toc504330748)

[4.1 Εισαγωγή 9](#_Toc504330749)

[4.2 Εκτέλεση 9](#_Toc504330750)

[4.2.1 Λειτουργία LL(1) Parser 12](#_Toc504330751)

[4.2.2 Εκτέλεση προγράμματος 15](#_Toc504330752)

[4.2.3 Στιγμιότυπα από την εκτέλεση του προγράμματος 15](#_Toc504330753)

[5 Άσκηση Δ 19](#_Toc504330754)

[5.1 Εισαγωγή 19](#_Toc504330755)

[5.2 Συντακτικό Διάγραμμα 19](#_Toc504330756)

[5.3 Περιγραφή EBNF 21](#_Toc504330757)

[5.4 Πρόγραμμα FLEX 21](#_Toc504330758)

[5.4.1 Στιγμιότυπα από την εκτέλεση του προγράμματος 22](#_Toc504330759)

[6 Άσκηση Ε 26](#_Toc504330760)

[6.1 Εισαγωγή 26](#_Toc504330761)

[6.2 Πρόγραμμα FLEX 27](#_Toc504330762)

[6.2.1 Στιγμιότυπα από την εκτέλεση του προγράμματος 27](#_Toc504330763)

# Άσκηση Α

## Εισαγωγή

Η πρώτη άσκηση έχεις ως στόχο την υλοποίηση ενός προγράμματος που δέχεται ως είσοδο ένα αρχείο πηγαίου κώδικα και μέσω μιας συγκεκριμένης διαδικασίας αναγνωρίζει τον αριθμό των παρενθέσεων και αποφαίνεται για την ορθότητά τους.

Το πρόγραμμα σε αρχικό στάδιο δέχεται ως όρισμα τη διαδρομή του αρχείο που θα λάβει ως είσοδο ώστε να αρχίσει την επεξεργασία. Έπειτα, διαβάζει τα περιεχόμενά του και αφαιρεί όλους τους χαρακτήρες εκτός των παρενθέσεων. Το πρόβλημα που καλείται το πρόγραμμα να αντιμετωπίσει έγκειται στον τρόπο με τον οποίο θα αναγνωριστούν οι ανοιχτές και κλειστές παρενθέσεις λαμβάνοντας υπ’ όψη τη σειρά με την οποία βρίσκονται.

## Επίλυση

Με τη χρήση ενός Ντετερμινιστικού Αυτόματου Στοίβας (ΝΑΣ) το ζήτημα αντιμετωπίζεται αρκετά πιο εύκολα. Σύμφωνα με τον ορισμό ενός ΝΑΣ, στη δική μας περίπτωση ορίζεται ως εξής:

**Μ = (K, T, V, p, k1, A1, F)**

**Όπου:**

Σύνολο Καταστάσεων (K)  
{ **k0**, **k1**, **k2** }

Αλφάβητο Εισόδου (T)  
{ **(**, **)** }

Αλφάβητο Συμβόλων (V)  
{ **(**, **)**, **$** }

Συνάρτηση Στοίβας (p)  
f(**k0**, **ε**, **ε**) = { (**k1**, **$**) }  
f(**k1**, **ε**, **(**) = { (**k1**, **(**) }  
f(**k1**, **(**, **)**) = { (**k1**, **ε**) }  
f(**k1**, **$**, **ε**) = { (**k2**, **ε**) }

Αρχική Κατάσταση (k1)  
**k0**

Αρχικό Σύμβολο Στοίβας (A1)  
**$**

Σύνολο Τελικών Καταστάσεων (F)  
{ **t2** }

Με άλλα λόγια, κατά την έναρξη του προγράμματος, αφού διαβαστούν τα περιεχόμενα του αρχείου, η στοίβα σύμφωνα με τη συνάρτησή της, αρχικοποιείται και μεταβαίνει στην κατάσταση **k1** με το σύμβολο **$** στο πυθμένα της στοίβας. Στη συνέχεια, εάν δεχτεί η στοίβα το σύμβολο **(** εισάγει τη παρένθεση και παραμένει στην κατάσταση **k1** μέχρις ότου το σύμβολο **)** να διαβαστεί. Σε αυτή τη περίπτωση, αφαιρείται η προηγούμενη παρένθεση **(** από τη στοίβα. Συνεχίζουμε αυτή τη διαδικασία μέχρι να καταναλωθεί η συμβολοσειρά από παρενθέσεις και να μεταβούμε στη τελική κατάσταση **k2**.

Κατά την κατανάλωση της συμβολοσειράς, εάν διαπιστωθεί πως δεν υπάρχει καθορισμένη μετάβαση σύμφωνα με τη συνάρτηση της στοίβας, η διαδικασία σταματάει και οι παρενθέσεις δεν αναγνωρίζονται.

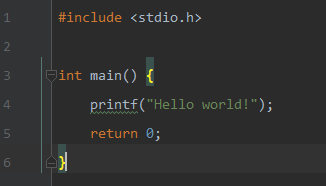
## Εκτέλεση

Το πρόγραμμα έχει υλοποιηθεί στη **Python 3** και δέχεται **1** όρισμα.

Ενδεικτικά, μέσω της γραμμής εντολών το πρόγραμμα εκτελείται ως εξής:

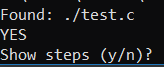
* Κάνουμε **cd compilers-project**
* Κάνουμε **cd exercise-1**
* Εκτελούμε το πρόγραμμα **python parentheses.py ./test.c**

Αν υποθέσουμε ότι το αρχείο **test.c** περιέχει τον πηγαίο κώδικα για ένα από hello world πρόγραμμα στη C και βρίσκεται στον ίδιο φάκελο με το αρχείο **parentheses.py**:



Εικόνα 2.1

Τότε παρατηρούμε ότι το πρόγραμμά μας ανταποκρίνεται ως εξής:

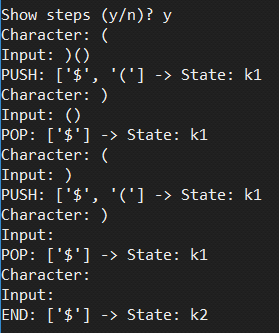


Εικόνα 2.2

Το πρόγραμμα βρήκε επιτυχώς το αρχείο **test.c** και αφού ανέλυσε τα περιεχόμενα και τις παρενθέσεις, τυπώνει «**YES**» στη γραμμή εντολών. Η ένδειξη αυτή σημαίνει πως οι παρενθέσεις που βρίσκονται στο πηγαίο κώδικα του test.c είναι σωστές.

Πράγματι αφού έχουμε τις εξής παρενθέσεις: **()()**

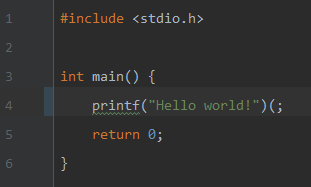
Στη συνέχεια, παρατηρούμε πως το πρόγραμμα δίνει την επιλογή στο χρήστη να του εκτυπώσει αναλυτικά όλα τα βήματα που οδήγησαν σε αυτό το αποτέλεσμα.



Εικόνα 2.3

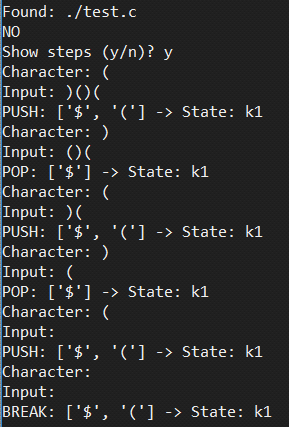
Στο τελευταίο βήμα, μπορούμε να διακρίνουμε πως η στοίβα περιέχει το σύμβολο **$** και βρίσκεται στην κατάσταση **k2**, άρα η συμβολοσειρά αναγνωρίζεται.

Σε μια άλλη περίπτωση που οι παρενθέσεις στο αρχείο του πηγαίου κώδικα δεν ήταν ισορροπημένες:



Εικόνα 2.4

Το αποτέλεσμα θα ήταν «**ΝΟ**» αφού πράγματι θα είχαμε τις εξής παρενθέσεις: **()()(**



Εικόνα 2.5

Έτσι, διαπιστώνουμε ότι η στοίβα παρέμεινε στη κατάσταση **k1** με μια εκκρεμής ανοιχτή παρένθεση.

# Άσκηση Β

## Εισαγωγή

Η δεύτερη άσκηση έχει ως θέμα την υλοποίηση ενός προγράμματος-γεννήτριας συμβολοσειρών σύμφωνα με την γραμματική που ορίζεται στην Ενότητα 3.2.4 του βιβλίου. Το πρόγραμμα όπως και στη πρώτη άσκηση έχει υλοποιηθεί στη γλώσσα προγραμματισμού Python, εφόσον είναι αρκετά ευέλικτη σε τέτοια θέματα παραγωγής συμβολοσειρών.

## Επίλυση

Ορίζουμε μια γραμματική G σύμφωνα με το παράδειγμα του βιβλίου.

**G = (S, P, N, T)**

Όπου:

Σύμβολο Εκκίνησης (S)  
**<έκφραση>**

Σύνολο Κανόνων Παραγωγής (P)  
**<έκφραση>** ::= <όρος> | <έκφραση> + <όρος>  
**<όρος>** ::= <παράγοντας> | <όρος> \* <παράγοντας>  
**<παράγοντας>** ::= a | b | c

Σύνολο μη Τερματικών Συμβόλων (N)  
{ **<έκφραση>**, **<όρος>**, **<παράγοντας>** }

Σύνολο Τερματικών Συμβόλων (Τ)  
{ **a**, **b**, **c** }

Η γραμματική G έχει οριστεί με BNF (Backus-Naur Form), μεταγλώσσα που χρησιμοποιείται για τον ορισμό γλωσσών χωρίς συμφραζόμενα. Η παρούσα γραμματική δεν έχει συμφραζόμενα, γι’ αυτό με τη χρήση BNF μπορούμε εύκολα να υλοποιήσουμε ένα πρόγραμμα για τη προτάσεων αυτής της γλώσσας.

Το πρόγραμμα σε αρχικό στάδιο, αρχίζει τη παραγωγή μιας πρότασης από την **<έκφραση>**. Έπειτα, επιλέγει τυχαία και εφαρμόζει τον κατάλληλο κανόνα για την αντικατάσταση του μη τερματικού συμβόλου. Η διαδικασία αυτή επαναλαμβάνεται μέχρις ότου να μην υπάρχει η <έκφραση> στη πρότασή μας. Η ίδια διαδικασία εφαρμόζεται και για τα σύμβολα **<όρος>** και **<παράγοντας>**. Όσο υπάρχουν μη τερματικά σύμβολα, επαναλαμβάνεται η εφαρμογή των κανόνων της γραμματικής.

Επειδή η εφαρμογή των κανόνων είναι τυχαία, μπορεί να καταλήξουμε σε μεγάλους βρόχους επανάληψης, δηλαδή σε μεγάλες προτάσεις που προέκυψαν έπειτα από την εφαρμογή πολλών κανόνων στη σειρά. Γι’ αυτό το λόγο, έχει τεθεί ένα όριο στον αριθμό των κανόνων που εφαρμόζονται. Έτσι αν έχουμε καταλήξει σε πρόταση που περιέχει μη τερματικά σύμβολα έπειτα από 50 εφαρμογές κανόνων, το πρόγραμμα σταματάει τη διαδικασία και εφαρμόζει τους πιο σύντομους κανόνες που θα μας οδηγήσουν σε τελικό αποτέλεσμα.

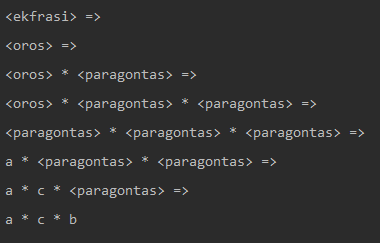
## Εκτέλεση

Το πρόγραμμα έχει υλοποιηθεί στη **Python 3** και δεν δέχεται ορίσματα.

Ενδεικτικά, μέσω της γραμμής εντολών το πρόγραμμα εκτελείται ως εξής:

* Κάνουμε **cd compilers-project**
* Κάνουμε **cd exercise-2**
* Εκτελούμε το πρόγραμμα **python generator.py**

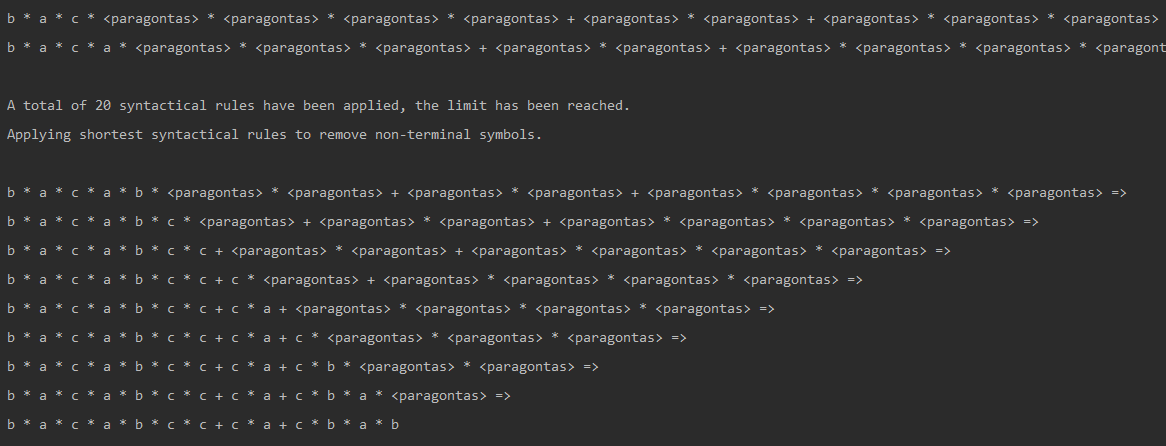
Αμέσως, με την εκτέλεση του προγράμματος παρατηρούμε πως έχει δημιουργηθεί μια πρόταση έπειτα από την εφαρμογή των κανόνων γραμματικής:



Εικόνα 3.1

Σε αυτή τη περίπτωση, το πρόγραμμα επέλεξε να αντικαταστήσει την **<έκφραση>** με **<όρος>**, έπειτα **<όρος>** με **<όρος> \* <παράγοντας>** κ.ο.κ. Έτσι, φτάσαμε στη συμβολοσειρά **a \* c \* b**.

Υπάρχουν και άλλες περιπτώσεις όπου το πρόγραμμα μπορεί να συνεχίσει τη εφαρμογή αναδρομικών κανόνων αρκετές φορές:



Εικόνα 3.2

Γι’ αυτό το λόγο, θέτουμε ένα όριο στην εφαρμογή των κανόνων ώστε αν έχουμε μεγάλους βρόχους επανάληψης, σταματάει η διαδικασία και καταλήγουμε όσο το συντομότερο δυνατό σε επιθυμητό αποτέλεσμα.

Στο παράδειγμά μας, το όριο έχει τεθεί σε 20 εφαρμογές κανόνων για τη παρουσίαση του παραδείγματος, όμως ρεαλιστικά το πρόγραμμα σταματάει μετά από **50 εφαρμογές**, που είναι αρκετά πιο σπάνιο.

# Άσκηση Γ

## Εισαγωγή

Το θέμα της τρίτης άσκησης είναι να εξετάσουμε αν η παρακάτω γραμματική είναι LL(1).

**S [A]**

**A BE**

**B x | y | S**

**E :A | +A | ε**

Επιπλέον μας ζητήθηκε να κατασκευάσουμε ένα συντακτικό αναλυτή top-down που να αναγνωρίζει την εκάστοτε δοθείσα συμβολοσειρά ή να απαντά αρνητικά ως προς την συντακτική της ορθότητα. Η γλώσσα προγραμματισμού η οποία επιλέχθηκε είναι η Java.

## Εκτέλεση

Αρχικά θα δείξουμε ότι η γραμματική είναι LL(1).

**FIRST**

FIRST(S) = {[}

FIRST(A) = {x, y, [}

FIRST(B) = {x, y, [}

FIRST(E) = {:, +, ε}

**FOLLOW**

1. **$ FOLLOW(S)**
2. **S [A]**

S [A]  **FIRST(]) – {ε} FOLLOW(A) = {]} FOLLOW(A)**

1. **A BE**

**A εΒΕ FIRST(E) – {ε} FOLLOW(B) = {:, +} FOLLOW(B)**

**A εΒΕ όπου ε FIRST(E) FOLLOW(A) FOLLOW(B)**

**A ΒΕ FOLLOW(A) FOLLOW(E)**

1. **E :A**

**Ε :Α FOLLOW(E) FOLLOW(A)**

1. **E +A**

**Ε +Α FOLLOW(E) FOLLOW(A)**

1. **B S**

**B S FOLLOW(B) FOLLOW(S)**

Επομένως έχουμε:

FOLLOW(Α) = {]} FOLLOW(Ε) = {]}

FOLLOW(B) = {:, +, ]} FOLLOW(S) = {$, :, +, ]}

**EMPTY**

EMPTY([A]) = FALSE

EMPTY(x) = FALSE

EMPTY(y) = FALSE

EMPTY(S) = FALSE

EMPTY(BE) = EMPTY(B) AND EMPTY(E) = FALSE AND TRUE = FALSE

EMPTY(:A) = EMPTY(:) AND EMPTY(A) = FALSE

EMPTY(+A) = EMPTY(+) AND EMPTY(A) = FALSE

EMPTY(ε) = TRUE

**LOOKAHEAD**

LOOKAHEAD(S [A]) = FIRST([) FIRST(A]) = {[, x, y}

LOOKAHEAD(**A ΒΕ**) = FIRST(B) FIRST(E) = {[, x, y, :, +, ]}

LOOKAHEAD(**B x**) = FIRST(x) = {x}

LOOKAHEAD(**B y**) = FIRST(y) = {y}

LOOKAHEAD(**B S**) = FIRST(S) = {[}

LOOKAHEAD(**Ε :Α**) = FIRST(:A) = {:}

LOOKAHEAD(**Ε +Α**) = FIRST(+A) = {+}

LOOKAHEAD(**Ε ε**) = FOLLOW(E) = { [}

Αρά έχουμε:

LOOKAHEAD(**B x**) LOOKAHEAD(**B y**) LOOKAHEAD(**B S**) =

LOOKAHEAD(**Ε :Α**) LOOKAHEAD(**Ε +Α**) LOOKAHEAD(**Ε +Α**) =

***Επομένως η γραμματική που δόθηκε είναι LL(1)***

Στη συνέχεια θα κατασκευάσουμε τον συντακτικό πίνακα, ο οποίος θα χρησιμοποιηθεί για την κατασκευή ενός αποδοτικού συντακτικού αναλυτή από πάνω προς τα κάτω.

1. **S [A]**

FIRST([) = [

Άρα:

Μ(S,[) = S [A]

1. **A ΒΕ**

FIRST(BE) = FIRST(B) = {x, y, [}

Άρα:

Μ(Α,x) = **A ΒΕ**

Μ(Α,y) = **A ΒΕ**

Μ(Α,[) = **A ΒΕ**

1. **B x | y | S**
   1. **B x**

**FIRST(x) = {x} Άρα Μ(Β, x) = B x**

* 1. **B y**

**FIRST(y) = {y] Άρα Μ(Β, y) = B y**

* 1. **B S**

**FIRST(S) = {[} Άρα Μ(Β, [) = B S**

1. **E :A | +A | ε**
   1. **Ε :Α**

**FIRST(:) = {:} Άρα Μ(E, :) = Ε :Α**

* 1. **Ε +Α**

**FIRST(+) = {+} Άρα Μ(Β, +) = Ε +Α**

* 1. **Ε ε**

**FIRST(ε) = {ε} ε FIRST(ε)**

**Αρά εξετάζουμε το FOLLOW(E) = {]}**

**Οπότε Μ(E, ]) = Ε ε**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| V/T | [ | ] | : | + | x | y | $ |
| S | S [A] |  |  |  |  |  |  |
| A | **A ΒΕ** |  |  |  | **A ΒΕ** | **A ΒΕ** |  |
| B | **B S** |  |  |  | **B x** | **B y** |  |
| E |  | **Ε ε** | **Ε :Α** | **Ε +Α** |  |  |  |

Πίνακας 1 Συντακτικός Πίνακας

### Λειτουργία LL(1) Parser

Ο LL(1) parser είναι υλοποιημένος σε γλώσσα προγραμματισμού JAVA. Συγκεκριμένα υλοποιήσαμε μια τάξη με όνομα LL1\_Parser και μέσα σε αυτή είναι περιέχονται οι μέθοδοι με τις απαραίτητες λειτουργίες. Το αρχείο java με τον κώδικα του συντακτικού αναλυτή βρίσκεται μέσα στο φάκελο **exercise-3/src**

**Μεταβλητές**

* **private String input**

Είναι private μεταβλητή τύπου string στην οποία γίνεται ανάθεση της συμβολοσειράς που είναι προς αναγνώριση.

* **private int indexOfInput**

Είναι private ακέραια μεταβλητή και χρησιμοποιείται για να κρατάει κάθε φορά το index του input στο οποίο βρισκόμαστε. Η αρχική της τιμή είναι -1.

* **private Stack<String> pareserStack**

Εδώ έχουμε μια private μεταβλητή η οποία έχει το ρόλο της στοίβας και δέχεται string.

* **String[][]** **syntaxArray**

Η μεταβλητή syntax Array αποτελεί ένα δυσδιάστατο πίνακα ο οποίος αντιπροσωπεύει τον συντακτικό πίνακα που κατασκευάσαμε παραπάνω.

* **private String[] termSymbols**

Είναι private μεταβλητή, με την οποία υλοποιούμε ένα πίνακα όπου δέχεται string. Ο συγκείμενος πίνακας περιέχει τα τερματικά σύμβολα.

* **private String[] nonTermSymbols**

Είναι private μεταβλητή, με την οποία υλοποιούμε ένα πίνακα όπου δέχεται string. Ο συγκείμενος πίνακας περιέχει τα μη τερματικά σύμβολα.

**Μέθοδοι**

* **public LL1\_Parser(String input)**

Αποτελεί τον κατασκευαστή της τάξης και δέχεται μία παράμετρο τύπου string. Έτσι όταν δημιουργείται ένα αντικείμενο της τάξης LL1\_Parser, μέσω της παραμέτρου αυτής, μπορούμε να περάσουμε την συμβολοσειρά στον Parser μας. Επιπλέον γίνεται και η δημιουργία της στοίβας.

* **public void parsingAlgorithm()**

Καλώντας αυτή τη μέθοδο, η οποία είναι public και τύπου void, ξεκινάει η διαδικασία της ανάλυσης.   
Μέσα σε αυτή περιέχεται όλη η βασική λειτουργία του parser.

Αρχικά εισάγεται στη στοίβα, κάτω, το σύμβολο $ και στη κορυφή το αρχικό σύμβολο S. Ο συντακτικός αναλυτής διαβάζει τα σύμβολα από την είσοδο ένα ένα.

Οι ενέργειες του αναλυτή καθορίζονται από το τρέχον σύμβολο εισόδου (*έστω a*) και το σύμβολο οπ βρίσκεται στη κορυφή της στοίβας (*έστω x*).

Υπάρχουν 4 περιπτώσεις για τον συντακτικό αναλυτή:

1. Εάν το x = a και a $ τότε το σύμβολο εισόδου είχε αναγνωριστεί και το x βγαίνει από την στοίβα.
2. Εάν x = a = $ τότε ο αναλυτής έχει δεχθεί τη συμβολοσειρά εισόδου και έχει τελειώσει την εργασία αναγνώρισης.

Εάν το x είναι μη τερματικό τότε ο αναλυτής ανατρέχει στο syntaxArray(x, a) του συντακτικού πίνακα.

1. Εάν στη θέση αυτή υπάρχει κανόνας παραγωγής τότε το x στη στοίβα αντικαθίσταται με τα σύμβολα στο δεξί μέλος της παραγωγής.

Τα σύμβολα μπαίνουν στη στοίβα με αντιστροφή σειρά. Δηλαδή το αριστερότερο σύμβολο παραγωγής καταλήγει στη κορυφή της λίστας.

1. Αν η θέση του πίνακα είναι κενή, τότε ο αναλυτής βγάζει μνήμα λάθους. Η συμβολοσειρά δεν αναγνωρίζεται από τη γραμματική σε αυτή την περίπτωση.

* **private void pushRuleInStack(String rule)**

Αυτή η μέθοδος είναι private τύπου void και δέχεται μια παράμετρο τύπου string. Η μέθοδος αυτή καλείται για την εισαγωγή του κανόνα στη στοίβα.

* **private boolean isTerminal(String inputString)**

Αυτή η μέθοδος είναι private, τύπου Boolean. Δέχεται μια παράμετρο τύπου string, η οποία ελέγχεται αν είναι τερματικό σύμβολο. Αν είναι η παράμετρος που δόθηκε είναι τερματικό σύμβολο επιστρέφεται η τιμή true, αλλιώς επιστρέφεται η τιμή false.

* **private String read()**

Η μέθοδος αυτή είναι private, τύπου string. Όταν καλείται αυξάνει τη μεταβλητή indexOfInput κατά ένα και στη συνέχεια επιστρέφει, από το Input, το χαρακτήρα που βρίσκεται στη συγκεκριμένη θέση.

* **private void errorLogger(String errorMsg)**

Η συγκεκριμένη μέθοδος είναι private, τύπου void, και στην περίπτωση που δεν αναγνωριστεί η συμβολοσειρά που δόθηκε εμφανίζει το κατάλληλο μήνυμα.

* **private String getRuleFromArray(String nonTerminal, String terminal)**

Η μέθοδος αυτή είναι private και τύπου string. Επιπλέον δέχεται δυο παραμέτρους τύπους string. Όταν καλείται, δίνονται ως πρώτη παράμετρος ένα μη τερματικό σύμβολο, και ως δεύτερη ένα τερματικό σύμβολο. Με βάση αυτά ελέγχεται αν υπάρχει ο κανόνας στο πίνακα και επιστρέφεται από τη μέθοδο. Αλλιώς εμφανίζει το κατάλληλο μήνυμα στο χρήστη.

* **private int getTermSymbolIndex(String inputString)**

Η μέθοδος αυτή είναι private, ακέραιου τύπου και δέχεται μια παράμετρο τύπου string. Στη συγκρίνεται η παράμετρος που δίνεται συγκρίνεται με τα στοιχεία που περιέχονται στον πίνακα με τα τερματικά σύμβολα. Αν η παράμετρος είναι ιδιά με κάποια από τα σύμβολα, επιστρέφεται το index στο οποίο βρέθηκε. Αλλιώς εμφανίζεται το κατάλληλο μήνυμα και επιστρέφεται η τιμή -1.

* **private int getNonTermSymbolIndex(String inputString)**

Η μέθοδος αυτή είναι private, ακέραιου τύπου και δέχεται μια παράμετρο τύπου string. Στη συγκρίνεται η παράμετρος που δίνεται συγκρίνεται με τα στοιχεία που περιέχονται στον πίνακα με τα μη τερματικά σύμβολα. Αν η παράμετρος είναι ιδιά με κάποια από τα σύμβολα, επιστρέφεται το index στο οποίο βρέθηκε. Αλλιώς εμφανίζεται το κατάλληλο μήνυμα και επιστρέφεται η τιμή -1.

* **private void printTableRow()**

Η μέθοδος αυτή είναι τύπου void και τυπώνει μια γραμμή για τον πίνακα αναγνώρισης της συμβολοσειράς.

* **private void printTableRow(string top, string rule, string token)**

Η μέθοδος αυτή είναι τύπου void, δέχεται 3 παραμέτρους, και τυπώνει μια γραμμή για τον πίνακα αναγνώρισης της συμβολοσειράς.

* **private String peekOfStack()**

Η μέθοδος αυτή είναι Private, τύπου String και επιστρέφει το στοιχείο που είναι στην κορυφή της στοίβας, χωρίς να το αφαιρεί από αυτήν.

* **private String popOutOfStack()**

Η μέθοδος αυτή είναι Private, τύπου String και επιστρέφει το στοιχείο που είναι στην κορυφή της στοίβας, αφαιρώντας το από αυτήν.

* **private void pushInStack(String string)**

Η μέθοδος αυτή είναι Private, τύπου void και δέχεται μια παράμετρο τύπου string. Η παράμετρος αυτή εισάγεται στη κορυφή της στοίβας.

### Εκτέλεση προγράμματος

Στον φάκελο exercise-3 υπάρχει και το αρχείο ll1\_parser.jar το οποίο είναι το εκτελέσιμο του προγράμματος. Για την εκτέλεση του ακολουθούμε τα παρακάτω βήματα:

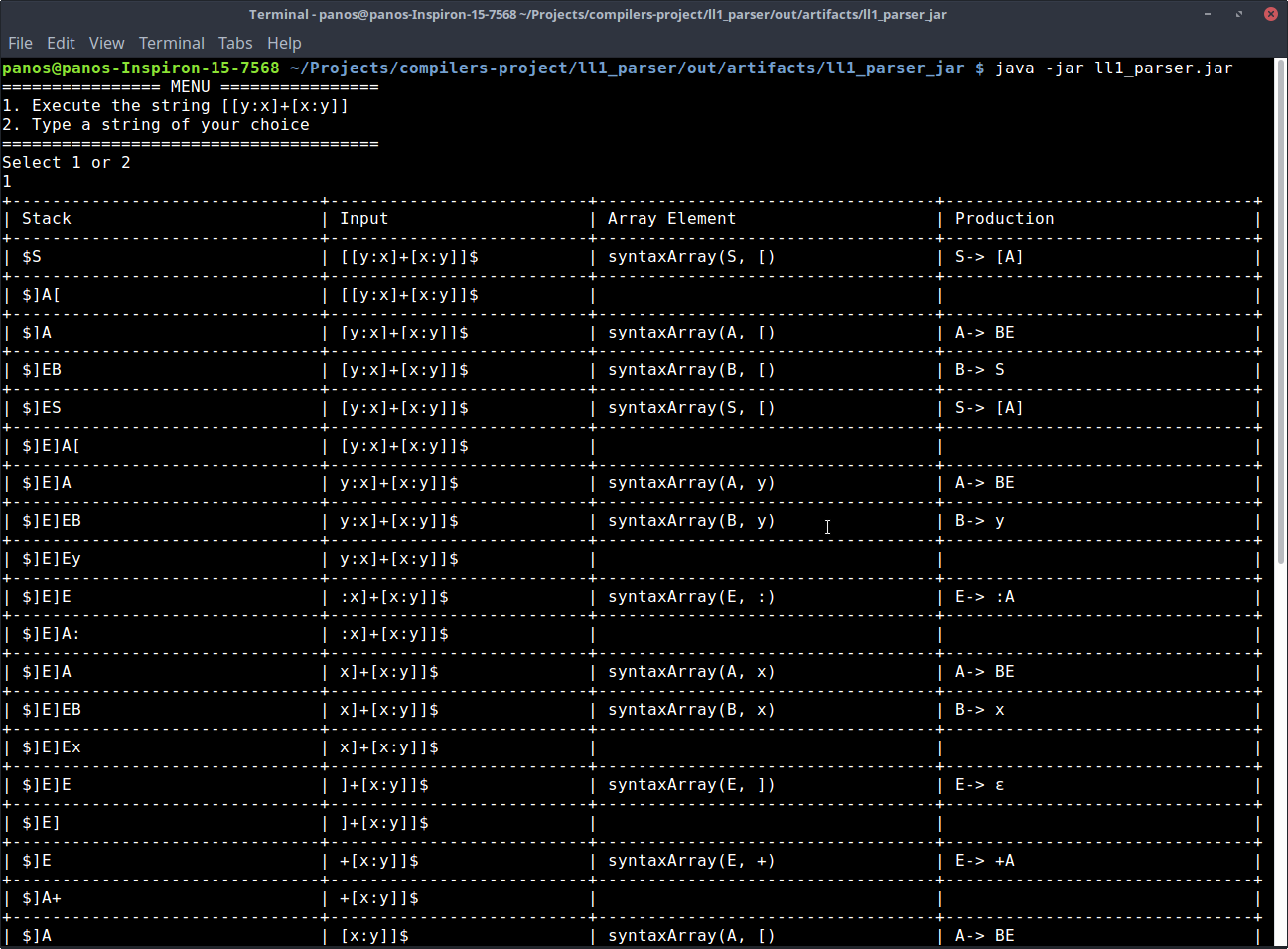
1. Ανοίγουμε το terminal και πάμε στο φάκελο **exercise-3** οπού περιέχεται το αρχείο **ll1\_parser.jar**
2. Γράφουμε την εντολή **java -jar ./ll1\_parser.jar**

### Στιγμιότυπα από την εκτέλεση του προγράμματος

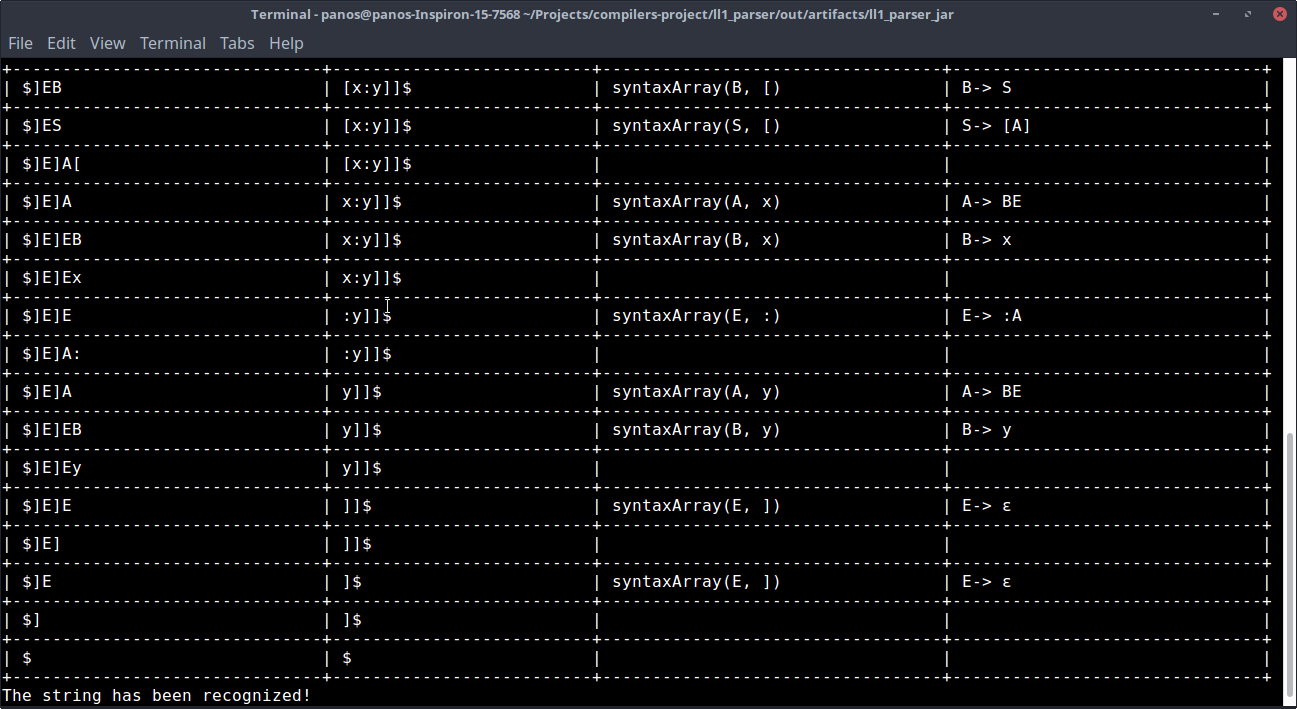
Παρακάτω ακολουθούν κάποια στιγμιότυπα από την εκτέλεση του προγράμματος. Με την εκτέλεση του προγράμματος εμφανίζεται ένα μενού στο οποίο ο χρήστης πρέπει να επιλέξει αναμεσά σε 2 επιλογές.

1. Να εκτελεστεί το πρόγραμμα έχοντας ως είσοδο την συμβολοσειρά του παραδείγματος.
2. Να εκτελεστεί το πρόγραμμα ζητώντας από τον χρήστη να εισάγει μια συμβολοσειρά της επιλογής του.

Αρχικά παρουσιάζεται η λειτουργία οπού το πρόγραμμα μας αναλύει την συμβολοσειρά του παραδείγματος, εικόνα 4.1 και εικόνα 4.2 (Η εικόνα 4.2 είναι συνέχεια της 4.1). Όπως φαίνεται η ανάλυση της συμβολοσειρά γίνεται βήμα - βήμα και εμφανίζεται ο αντίστοιχος πίνακας. Επιπλέον στο τέλος της εικόνας 4.2 μπορούμε να δούμε και το μήνυμα που εμφανίζεται, το οποίο ενημερώνει το χρήστη ότι η συμβολοσειρά αναγνωρίστηκε επιτυχώς.

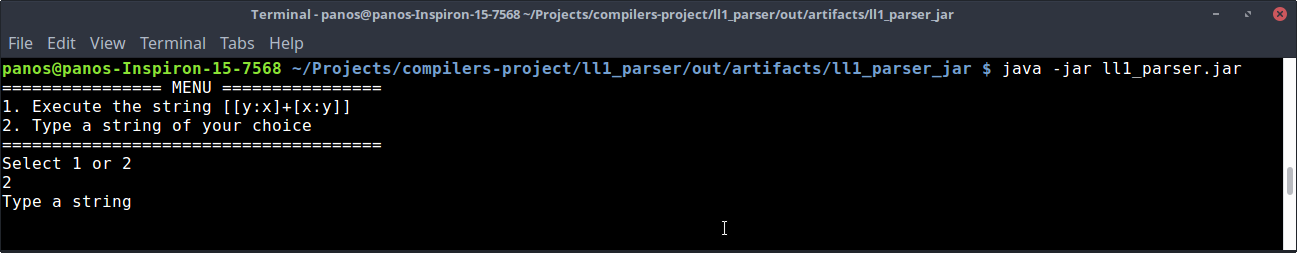


Εικόνα 4.1



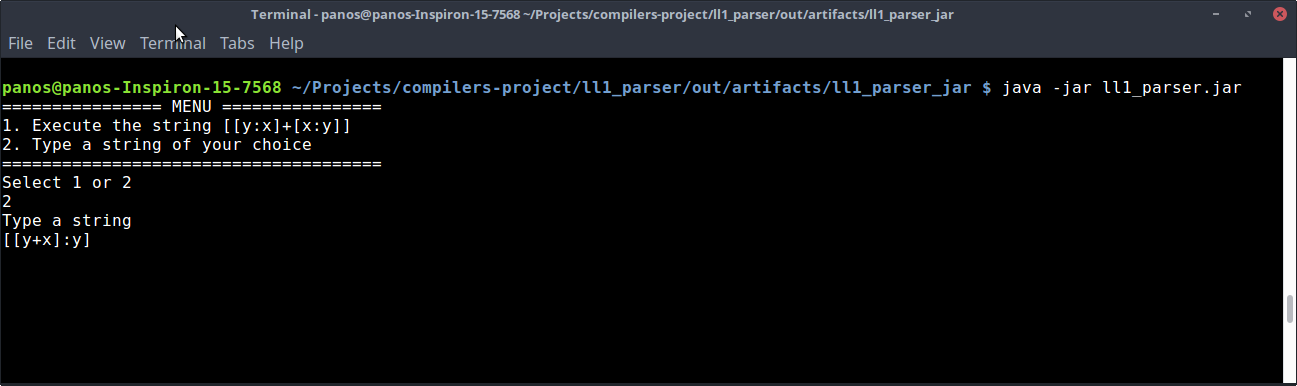
Εικόνα 4.2

Τώρα στην περίπτωση που ο χρήστης επιλέξει από το αρχικό μενού την επιλογή 2, θα πρέπει να εισάγει μια συμβολοσειρά, εικόνα 4.3.



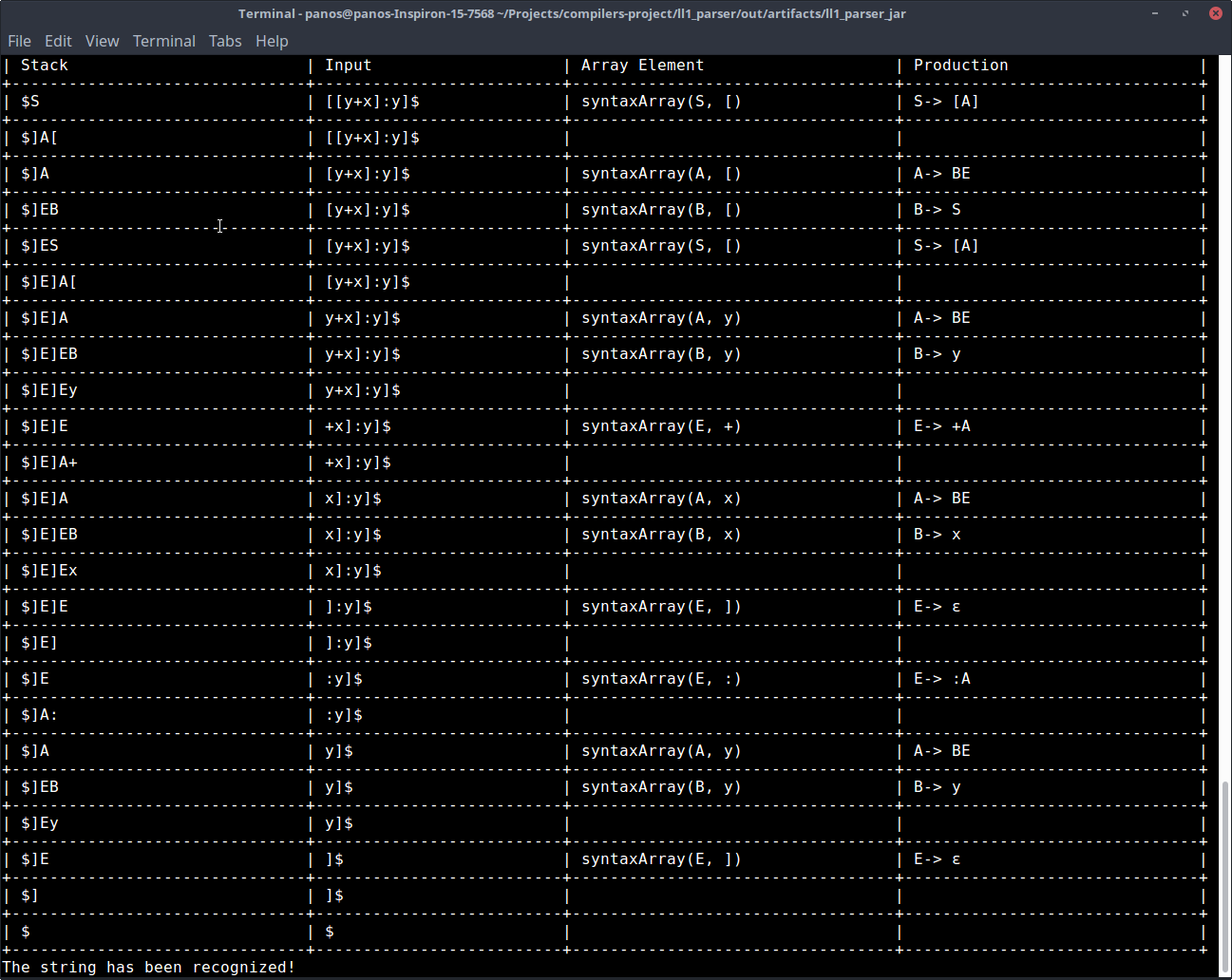
Εικόνα 4.3

Για τις ανάγκες της παρουσιάσης εισήγαμε την συμβολοσειρά [[y+x]:y], εικόνα 4.4.



Εικόνα 4.4

Στην εικόνα 4.5 φαίνεται ο πίνακας καθώς και το μήνυμα που ενημερώνει τον χρήστη ότι η συμβολοσειρά που εισήγαγε αναγνωρίστηκε επιτυχώς.



Εικόνα 4.5

Να επισημάνουμε ότι και σε περίπτωση που δεν αναγνωριστεί η συμβολοσειρά ή προκύψει κάποιο σφάλμα, πάντα ενημερώνεται ο χρήστης με αντίστοιχα μηνύματα.

# Άσκηση Δ

## Εισαγωγή

Σε αυτή την άσκηση μας δόθηκε το τμήμα δήλωσης μεταβλητών μια γλώσσας και μας ζητήθηκε να υλοποιήσουμε τα ακόλουθα:

1. Το συντακτικό διάγραμμα
2. Η περιγραφή EBNF
3. Το πρόγραμμα Flex

Ποιο συγκεκριμένα, στην αρχή θα πρέπει να βρίσκεται η λέξη «var». Στη συνέχεια τα ονόματα των μεταβλητών του ιδίου τύπου χωρισμένα με κόμματα. Αφού τελειώσει η παράθεση των ονομάτων, μπαίνει άνω-κάτω τελειά «:» και στη συνέχεια γράφεται ο τύπος των μεταβλητών. Στη συνέχεια δίνεται η δυνατότητα να οριστούν και άλλες μεταβλητές, ακολουθώντας πάντα την παραπάνω διαδικασία. Οι 4 τύποι των μεταβλητών που θα αναγνωρίζονται είναι: integer, Boolean, char, real. Ένα παράδειγμα που είναι έγκυρο για την γλώσσα που έχει περιγράφει είναι το ακόλουθο:

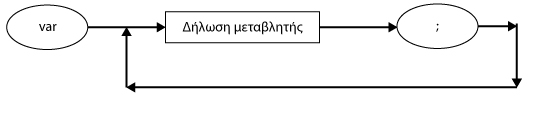
**var number\_of\_attendants, sum: integer;**

**ticket\_price: real;**

**symbols: char;**

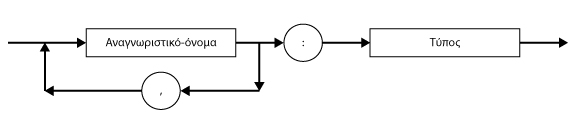
## Συντακτικό Διάγραμμα

Το σχήμα 5.1 δείχνει το συντακτικό διάγραμμα για τον ορισμό του τμήματος δηλώσεων.



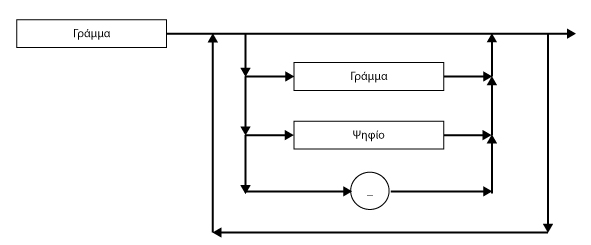
Σχήμα 5.1

Το σχήμα 5.2 δείχνει τη δήλωση της μεταβλητής.

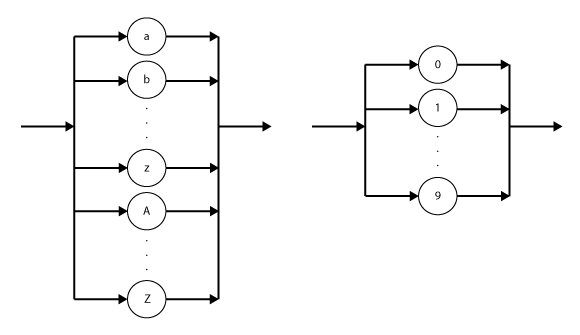


Σχήμα 2

Το σχήμα 5.3 δείχνει τον ορισμό του αναγνωριστικού ονόματος.



Σχήμα 5.3

Το σχήμα 5.4 δείχνει τον ορισμό του γράμματος και του ψηφίου.

Σχήμα 5.4

## Περιγραφή EBNF

**variable\_declaration = “var” variable\_list.**

**variable\_list = indentifier\_list “:” type.**

**identifier = letter{letter|digit|\_}.**

**letter = “A”|“B”|“C”|“D”|“E”|“F”|“G”|“H”|“I”|“J”|“K”|“L”|“M”|“N”| “O”|“P”|“Q”|“R”|“S”|“T”|“U”|“V”|“W”|“X”|“Y”|“Z”|“a”|“b”|“c”|“d”|“e”|“f”|“g”|“h”|“i”|“j”|“k”|“l”|“m”|“n”|“o”|“p”|“q”|“r”|“s”|“t”|“u”|“v”|“w”|“x”|“y”|“z”.**

**digit = “0”|“1”|“2”|“3”|“4”|“5”|“6”|“7”|“8”|“9”.**

**type = (“real”|“char”|“boolean”|“integer”)“;”[variable\_list].**

## Πρόγραμμα FLEX

Το πρόγραμμα μας υποστηρίζει και την εισαγωγή δηλώσεων μεταβλητών σε πολλαπλές γραμμές, για αυτό το λόγο πρέπει να πατήσουμε το πλήκτρο «enter» 2-3 ώστε το πρόγραμμα μας να καταλάβει ότι τελειώσαμε.

Όσο αυτά που εισάγουμε είναι έγκυρα, με το πάτημα του «enter» θα αλλάζουμε γραμμή. Αν πάλι αυτό που πληκτρολογήσαμε δεν είναι έγκυρο την στιγμή που πατήθηκε το «enter» για να αλλάξουμε γραμμή, το πρόγραμμα δεν θα μας αφήσει να, οπότε πρέπει να κάνουμε μια νέα εισαγωγή.

Το αρχείο που πρόγραμμα flex βρίσκεται στο φάκελο exercise-4 και έχει όνομα exercise4.l

Για να δημιουργήσουμε το εκτελέσιμο αρχείο ακολουθούμε τα παρακάτω βήματα.

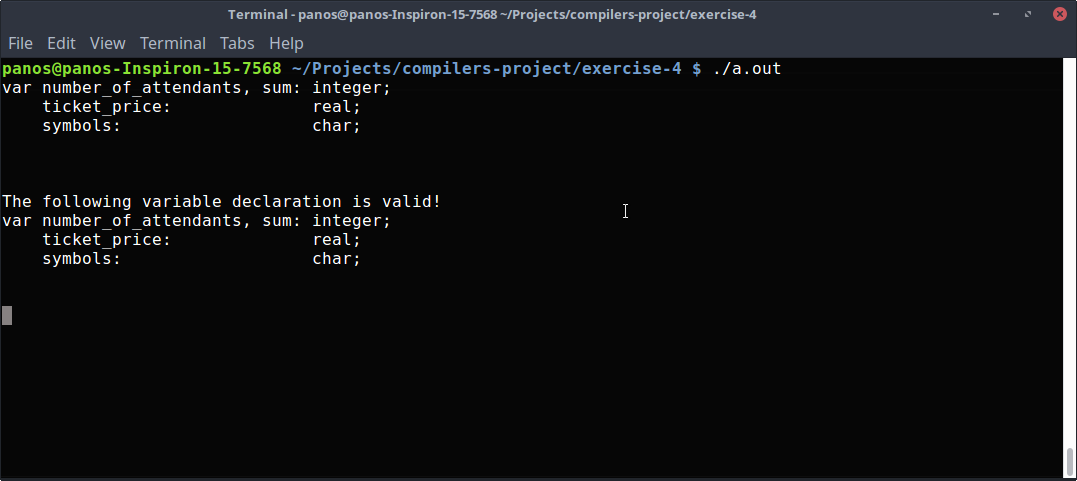
1. Ανοίγουμε το terminal και πάμε στο φάκελο **exercise-4** οπού περιέχεται το αρχείο **exercise4.l**
2. Γράφουμε την εντολή **flex exercise4.l**
3. Γράφουμε την εντολή **gcc lex.yy.c -lfl**

Εκτελούμε το αρχείο με την εντολή **./a.out**

### Στιγμιότυπα από την εκτέλεση του προγράμματος

Παρακάτω ακολουθούν στιγμιότυπα που παρουσιάζουν διάφορες φάσης της εκτέλεσης του προγράμματος.

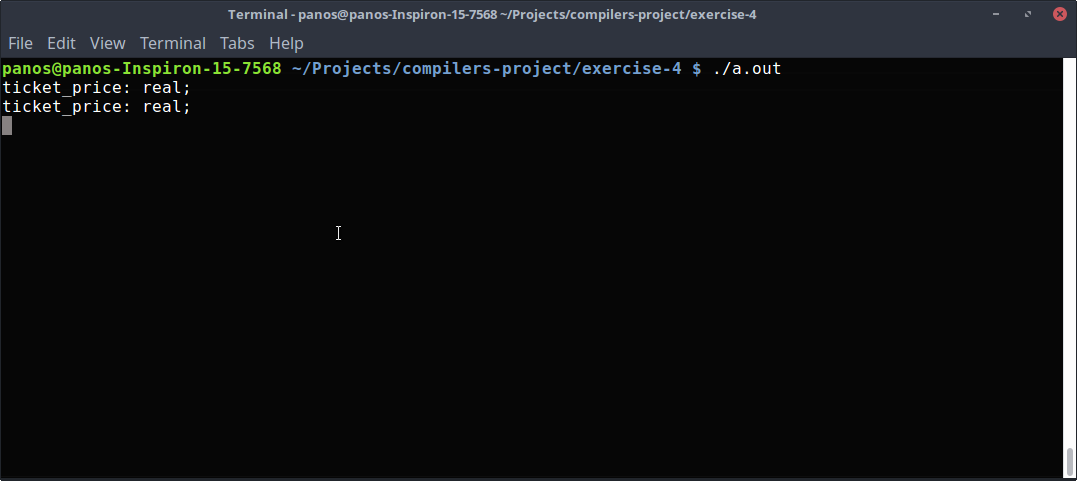
Στην εικόνα 6.1 δώσαμε ως είσοδο το παράδειγμα που είχαμε παραπάνω και όπως βλέπουμε αναγνωρίζεται κανονικά.



Εικόνα 6.2

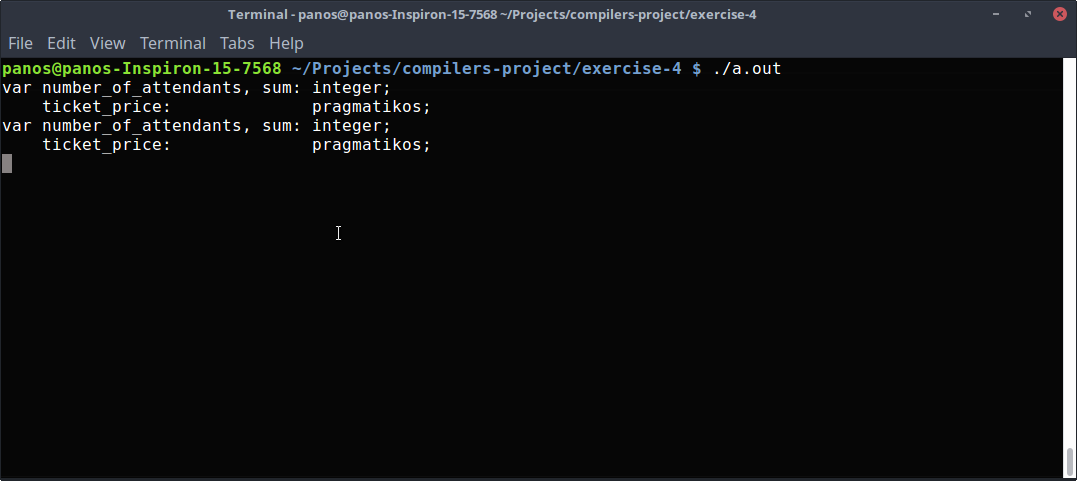
Στις επόμενες εικόνες παρουσιάζονται περιπτώσεις οπού η αναγνώριση δεν ήταν επιτυχής.

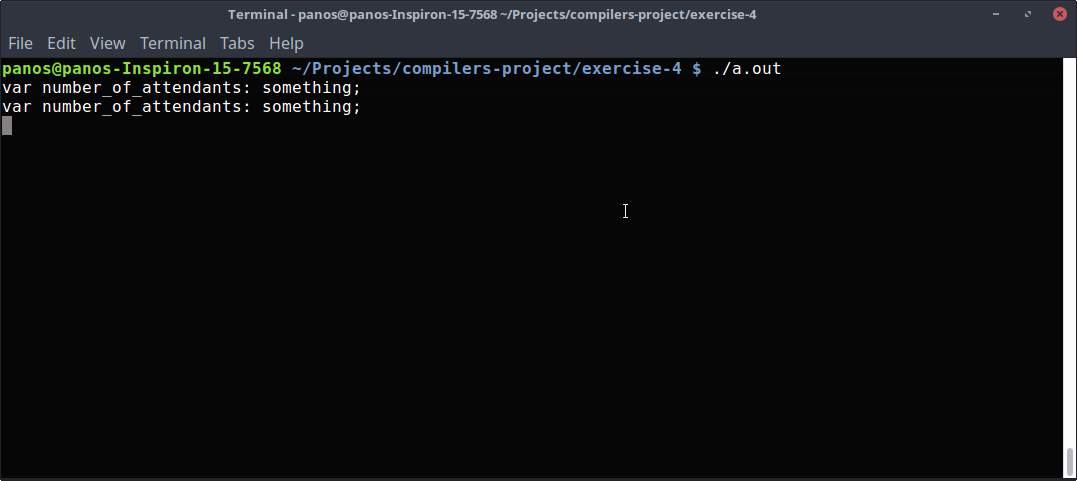
Στην εικόνα 6.2, όπως φαίνεται ξεκινάμε χωρίς την λέξη «var».



Εικόνα 6.3

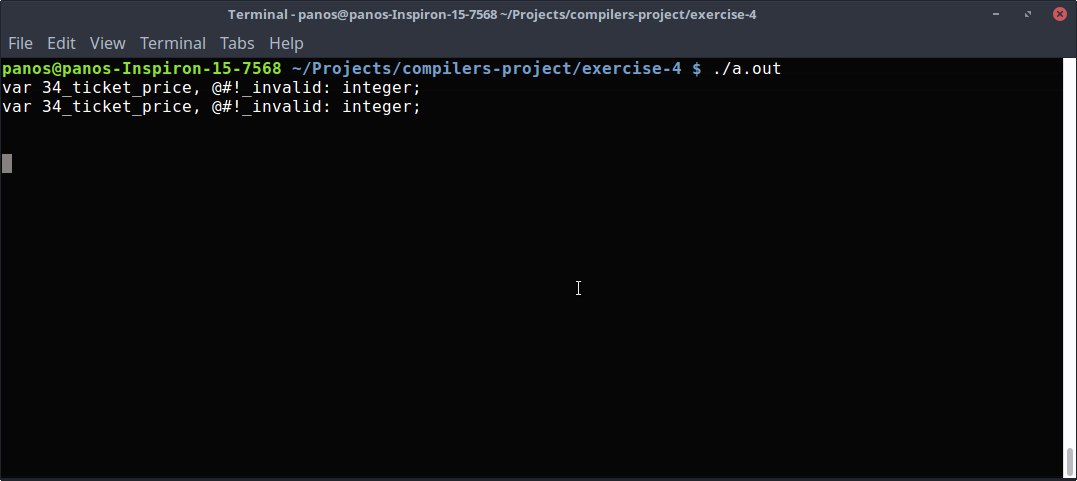
Στα επόμενα δυο παραδείγματα βλέπουμε τις περιπτώσεις που δίνεται ως τύπος κάποιο μη αποδεκτό όνομα. Στην εικόνα 6.3 βλέπουμε ότι αντί για «real» γράψαμε «πραγματικός», ενώ στην εικόνα 6.4 αντί για «integer» γράψαμε «something».



Εικόνα 6.3

Εικόνα 6.4

Στο τελευταίο μας παράδειγμα έχουμε δώσει μη αποδεκτά ονόματα μεταβλητών. Όπως φαίνεται στην εικόνα 6.5, το πρώτο όνομα (34\_ticket\_number) ξεκινάει με αριθμό κάτι το οποίο δεν είναι αποδεκτό. Επίσης και το δεύτερο όνομα δεν είναι αποδεκτό (@#!\_invalid) διότι ξεκινάει με σύμβολο.



Εικόνα 6.4

# Άσκηση Ε

## Εισαγωγή

Στη άσκηση αυτή μας ζητήθηκε να γράψουμε ένα πρόγραμμα FLEX το οποίο θα αναλύει προτάσεις της μορφής «Dinetai trigwno ABC». Έτσι λοιπόν στην προηγουμένη πρόταση θα πρέπει να λάβουμε ως έξοδο τα εξής:

***Dinetai: einai rhma***

***trigwno: einai gewmetrikh ontothta***

***ABC: einai onoma gewmetrikhs ontothtas***

Στο πρόγραμμα μας λοιπόν έχουμε υλοποιήσει τις περίπτωσης για την γωνιά, την ευθεία, το τρίγωνο, το τετράγωνο και το πεντάγωνο.

Επίσης, το πρόγραμμα είδη είναι δέχεται greeklish μπορεί να αναγνωρίζει της ακόλουθες παραλλαγές εκφράσεων:

* Για το «Dinetai» που είναι το ρήμα έχουμε την κανονική έκφραση

**(D|d)inetai** , ώστε να αναγνωρίζεται και το «dinetai» αλλά και το «Dinetai» , διότι βρίσκεται στην αρχή της πρότασης.

* Για το «euthia» που είναι γεωμετρική οντότητα έχουμε την κανονική έκφραση

**e(y|u)thia**, ώστε να αναγνωρίζεται και το «eythia» αλλά και το «euthia».

* Για το «trigwno» που είναι γεωμετρική οντότητα έχουμε την κανονική έκφραση **trig(w|o)no**, ώστε να αναγνωρίζεται και το «trigwno» αλλά και το «trigono».
* Για το «tetragwno» που είναι γεωμετρική οντότητα έχουμε την κανονική έκφραση **tetrag(w|o)no**, ώστε να αναγνωρίζεται και το «tetragwno» αλλά και το «tetragono».
* Για το «gwnia» που είναι γεωμετρική οντότητα έχουμε την κανονική έκφραση

**g(w|o)nia**, ώστε να αναγνωρίζεται και το «gwnia» αλλά και το «gonia».

* Για το «polygwno» που είναι γεωμετρική οντότητα έχουμε την κανονική έκφραση **pentag(w|o)no**, ώστε να αναγνωρίζεται και το «polygwno» αλλά και το «polygono».

Τέλος για τα 5 γράμματα της αλφάβητου (A, B, C, D, E) χρησιμοποιήσαμε την κανονική έκφραση [A-E]

Σε αυτό το σημείο να σημειώσουμε ότι έχει γίνει και η υλοποίηση του ερωτήματος Ε1 αλλά και του Ε2. Συνεπώς ο αναλυτής μας δεν θα αναγνώριση για παράδειγμα κάποια είσοδο της μορφής *«Dinetai trigwno AAB»* γιατί το όνομα της γεωμετρικής οντότητας περιέχει δυο φορές το γράμμα A. Επίσης δεν θα αναγνωρίσει για παράδειγμα την είσοδο *«Dinetai eythia AECD»* γιατί το όνομα της ευθείας πρέπει να αποτελείται μόνο από 2 γράμματα. Στη συνέχεια ακολουθών στιγμιότυπα από την εκτέλεση του προγράμματος, με διαφορά παραδείγματα.

## Πρόγραμμα FLEX

Το αρχείο που περιέχει το πρόγραμμα flex βρίσκεται στο φάκελο **exercise-5** και έχει όνομα **exercise5.l**

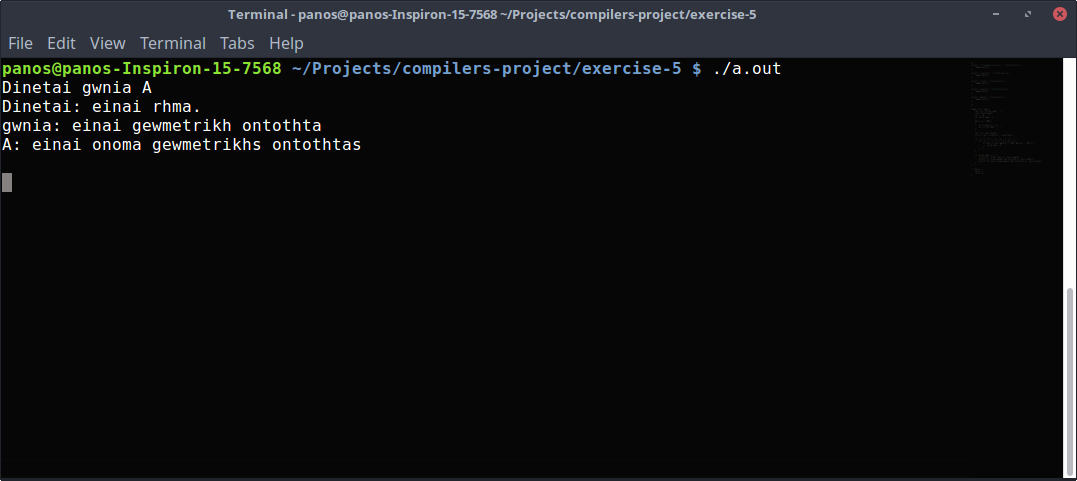
Για να δημιουργήσουμε το εκτελέσιμο αρχείο ακολουθούμε τα παρακάτω βήματα.

1. Ανοίγουμε το terminal και πάμε στο φάκελο **exercise-5** οπού περιέχεται το αρχείο **exercise5.l**
2. Γράφουμε την εντολή **flex exercise5.l**
3. Γράφουμε την εντολή **gcc lex.yy.c -lfl**
4. Εκτελούμε το αρχείο με την εντολή **./a.out**

### Στιγμιότυπα από την εκτέλεση του προγράμματος

Παρακάτω ακολουθούν μερικά παραδείγματα για κάθε περίπτωση, ώστε να δούμε πω λειτουργεί το πρόγραμμα.

#### Γωνία

Στην εικόνα 6.1, παρουσιάζεται μια έγκυρη εισαγωγή για τη γωνιά. Όπως φαίνεται αναγνωρίζεται κανονικά.

Εικόνα 6.1

Στην εικόνα 6.2 το όνομα της γεωμετρικής οντότητας από 2 γράμματα, κάτι το οποίο δεν είναι αποδεκτό.



Εικόνα 6.2

#### Ευθεία

Στην εικόνα 6.3, παρουσιάζεται μια έγκυρη εισαγωγή για τη ευθεία. Όπως φαίνεται αναγνωρίζεται κανονικά

Εικόνα 6.3

Στην εικόνα 6.4 το όνομα της γεωμετρικής οντότητας αποτελείται από 3 γράμματα, κάτι το οποίο δεν είναι αποδεκτό.



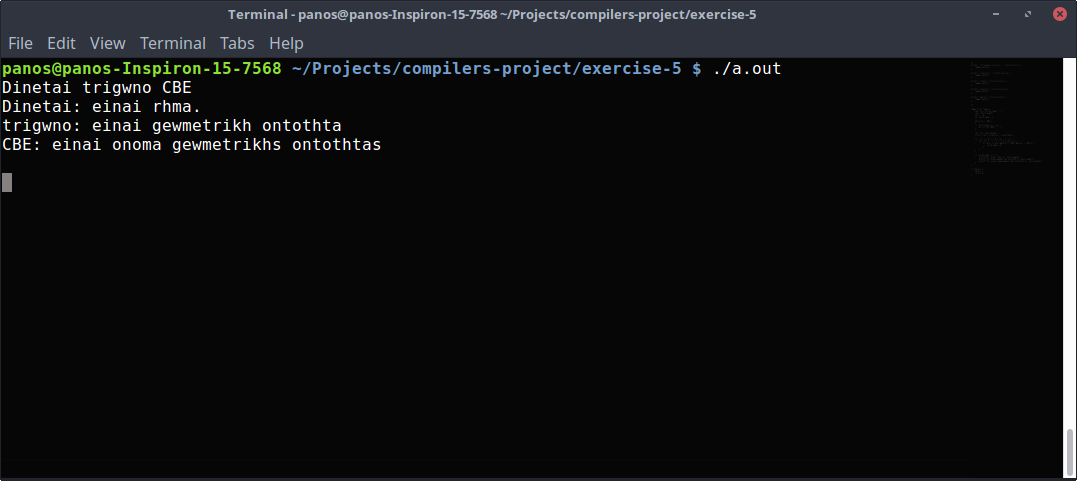
Εικόνα 6.4

Επίσης παρατηρούμε ότι δεν δέχεται το ίδιο γράμμα κατά τον ορισμό του ονόματος, εικόνα 6.5.



Εικόνα 6.5

#### Τρίγωνο

Στην εικόνα 6.6, παρουσιάζεται μια έγκυρη εισαγωγή για το τρίγωνο. Όπως φαίνεται αναγνωρίζεται κανονικά

Εικόνα 6.6

Στην εικόνα 6.7 το όνομα της γεωμετρικής οντότητας αποτελείται από 5 γράμματα, κάτι το οποίο δεν είναι αποδεκτό.



Εικόνα 6.7

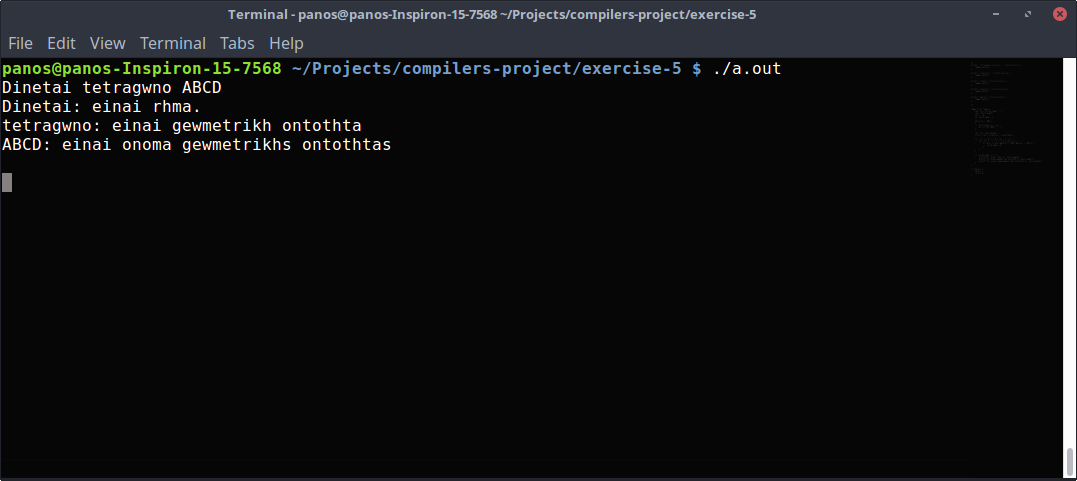
Επίσης παρατηρούμε ότι δεν δέχεται το ίδιο γράμμα κατά τον ορισμό του ονόματος, εικόνα 6.8.



Εικόνα 6.8

#### Τετράγωνο

Στην εικόνα 6.9, παρουσιάζεται μια έγκυρη εισαγωγή για το τετράγωνο. Όπως φαίνεται αναγνωρίζεται κανονικά



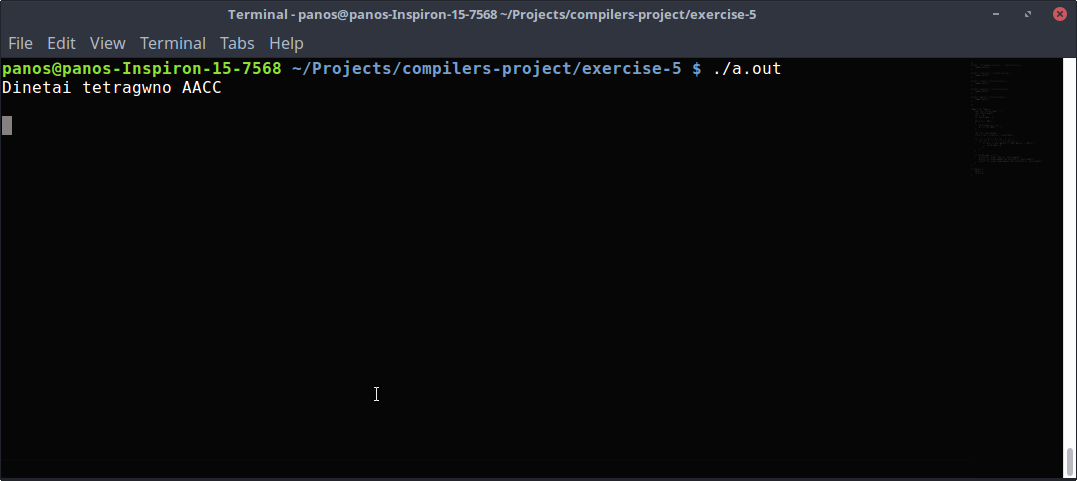
Εικόνα 6.9

Στην εικόνα 6.10 το όνομα της γεωμετρικής οντότητας αποτελείται από 2 γράμματα, κάτι το οποίο δεν είναι αποδεκτό.



Εικόνα 6.10

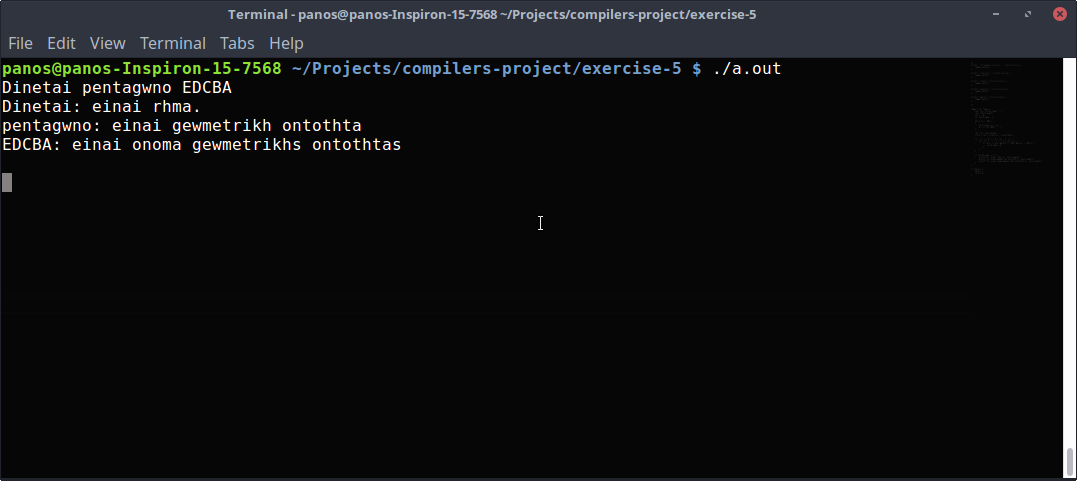
Επίσης παρατηρούμε ότι δεν δέχεται το ίδιο γράμμα κατά τον ορισμό του ονόματος, εικόνα 6.11.



Εικόνα 6.11

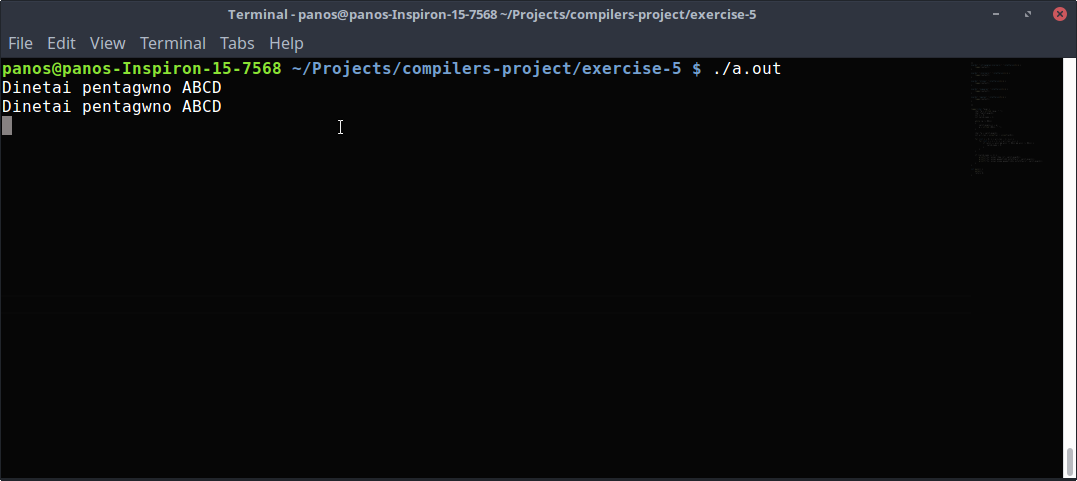
#### Πεντάγωνο

Στην εικόνα 6.12, παρουσιάζεται μια έγκυρη εισαγωγή για το πεντάγωνο. Όπως φαίνεται αναγνωρίζεται κανονικά



Εικόνα 6.12

Στην εικόνα 6.13 το όνομα της γεωμετρικής οντότητας αποτελείται από 4 γράμματα, κάτι το οποίο δεν είναι αποδεκτό.



Εικόνα 6.13

Επίσης παρατηρούμε ότι δεν δέχεται το ίδιο γράμμα κατά τον ορισμό του ονόματος, εικόνα 6.14.



Εικόνα 6.14