**ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ**

**ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ**

**ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ**

**Εικόνα που περιέχει clipart, θηλαστικό, γραφικά, εικονογράφηση

Περιγραφή που δημιουργήθηκε αυτόματα**

ΜΑΘΗΜΑ: ΜΕΤΑΓΛΩΤΤΙΣΤΕΣ

ΕΡΓΑΣΙΑ 1η ΜΕΡΗ Α2

ΤΜΗΜΑ: Β2

ΟΜΑΔΑ: 1η

ΜΕΛΗ ΟΜΑΔΑΣ:

1. Φράγκος Μαρίνος (ΠΑΔΑ – 20390255)
2. Φριλίγκος Γρηγόριος (ΠΑΔΑ – 20390258)
3. Βροχάρης Αντώνιος (ΠΑΔΑ – 20390030)
4. Φάσσου Κοντοδημάκη Ιφιγένεια Γεωργία (ΠΑΔΑ – 20390249)
5. Διάννης Ιωάννης (ΠΑΔΑ – 21390053)

Πίνακας περιεχομένων

[1. Εισαγωγή 3](#_Toc164779087)

[2. Λεκτικές Μονάδες 4](#_Toc164779088)

[2.1 Αναγνωριστικά: 4](#_Toc164779089)

[2.1.1 Αυτόματο 5](#_Toc164779090)

[2.1.2 Πίνακας Μεταβάσεων Αυτόματου 5](#_Toc164779091)

[2.1.3 Κώδικας FSM 5](#_Toc164779092)

[2.2 Λέξεις κλειδιά: 6](#_Toc164779093)

[2.3 Λεκτικά Κυριολεκτικά: 6](#_Toc164779094)

[2.3.1Αυτόματο 6](#_Toc164779095)

[2.3.2 Πίνακας Μεταβάσεων Αυτόματου 7](#_Toc164779096)

[2.3.3Κώδικας FSM 7](#_Toc164779097)

[2.4 Ακέραιοι: 7](#_Toc164779098)

[2.4.1 Αυτόματο 8](#_Toc164779099)

[2.4.2 Πίνακας Μεταβάσεων Αυτόματου 8](#_Toc164779100)

[2.4.3 Κώδικας FSM 8](#_Toc164779101)

[2.5 Αριθμοί Κινούμενης υποδιαστολής: 9](#_Toc164779102)

[2.5.1 Αυτόματο 10](#_Toc164779103)

[2.5.2 Πίνακας Μεταβάσεων Αυτόματου 10](#_Toc164779104)

[2.5.3 Κώδικας FSM 10](#_Toc164779105)

[2.6 Τελεστές: 11](#_Toc164779106)

[2.7 Σχόλια: 11](#_Toc164779107)

[2.7.1 Αυτόματο 12](#_Toc164779108)

[2.7.2 Πίνακας Μεταβάσεων Αυτόματου 12](#_Toc164779109)

[2.7.3 Κώδικας FSM 12](#_Toc164779110)

[2.8 White\_spaces χαρακτήρες: 13](#_Toc164779111)

[Ενιαίο Αυτόματο 14](#_Toc164779112)

[Αυτόματο 14](#_Toc164779113)

[Πίνακας Μεταβάσεων Αυτόματου 14](#_Toc164779114)

[Κώδικας FSM 14](#_Toc164779115)

# Εισαγωγή

Στην πρώτη εργαστηριακή άσκηση καλούμαστε να υλοποιήσουμε το μέρος Α-2 το οποίο ασχολείται με την κωδικοποίηση αυτόματων πεπερασμένων καταστάσεων FSM. Αρχικά, πρέπει να σχεδιάσουμε τις κανονικές εκφράσεις των λεκτικών μονάδων της γλώσσας καθώς και τα αντίστοιχα πεπερασμένα αυτόνομα συστήματα αναγνώρισής τους.

Στη συνέχεια, αφού τα δημιουργήσουμε, θα πρέπει να τα συνενώσουμε σε ένα ενιαίο αυτόματο σύστημα στο οποίο θα βασιστούμε ώστε να το προσομοιώσουμε σε έναν γενικό Πίνακα Μεταβάσεων, ο οποίος θα περιέχει τις διαδρομές των καταστάσεων που υπάρχουν στο ενιαίο αυτόματο.

Με τη χρήση του εργαλείου FSM , θα κωδικοποιήσουμε τον γενικό Πίνακα Μεταβάσεων με σκοπό να ελέγξουμε την ορθή αναγνώριση των λεκτικών μονάδων της γραμματικής. Σε αυτή τη διαδικασία της σύνταξης του κώδικα FSM, θα προχωρήσουμε σε κατάλληλες δοκιμές οι οποίες θα μας επιτρέψουν να ελέγξουμε την λειτουργικότητα και την αποτελεσματικότητα του ενιαίου αυτόματου.

Τέλος, στην εργασία θα αποδώσουμε ένα γενικό συμπέρασμα για το σύνολο των ενεργειών που υλοποιήσαμε καθώς και για τυχόν παραδοχές ή προβλήματα που συναντήσαμε στον κώδικα ή το ενιαίο αυτόματο.

# 2. Λεκτικές Μονάδες

Τα στοιχεία μιας γλώσσας διαχωρίζονται και όταν με βάση ένα πρότυπο αναγνώρισης

αναγνωριστούν επιτυχώς, χωρίζονται σε λεκτικές μονάδες. Αυτός ο χωρισμός σε λεκτικές μονάδες συχνά δεν είναι προκαθορισμένος. Σε περιπτώσεις όπου υπάρχει ασάφεια κατά το διαχωρισμό λέξεων, ο διαχωρισμός περιλαμβάνει την μακρύτερη πιθανή συμβολοσειρά στην οποία σχηματίζεται ένα αποδεκτό token όταν η συμβολοσειρά εισόδου διαβάζεται κατά το διαχωρισμό της από αριστερά προς τα δεξιά.

## 2.1 Αναγνωριστικά:

Ονόματα μεταβλητών που δέχονται χαρακτήρες [a-z] [A-Z] ,\_ και εκτός από τον πρώτο χαρακτήρα τους αριθμούς [0-9] με όριο μήκους 32 χαρακτήρων.

**Κώδικας σε Regex:**

A-Za-z\_][A-Za-z0-9\_]{0,32}

**Ενδεικτικά παραδείγματα από το περιβάλλον RegEx Pal:**

x

sum

\_value

myVariable55

\_\_counter

MAX\_SIZE

myFunction

3variable

!x

Αποδεκτά γίνονται όλα εκτός από το 3variable διότι ξεκινάει με αριθμό και το !x διότι ξεκινά με σύμβολο διαφορετικό του \_.

### 2.1.2 Πίνακας Μεταβάσεων Αυτόματου

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Χαρακτήρες Εισόδου | | | | | |
| Καταστάσεις |  | A-Z a-z \_ | 0-9 | \* | \n |
| ID | ID\_1 |  |  |  |
| ID\_1 | ID\_1 | ID\_1 |  | GOOD |

### 2.1.1 Αυτόματο

Εικόνα που περιέχει κύκλος, γραμμή, κείμενο, διάγραμμα

Περιγραφή που δημιουργήθηκε αυτόματα

### 2.1.3 Κώδικας FSM

START = ID

ID: A-Z a-z \_ -> ID\_1

\* -> BAD

ID\_1: A-Z a-z \_ 0-9 -> ID\_1

\n -> GOOD

\* -> BAD

GOOD(OK):

## 2.2 Λέξεις κλειδιά:

Σύμφωνα με την εκφώνηση όσο αφορά τις λέξεις κλειδιά δεν απαιτούνται κανονικές εκφράσεις Κ.Ε. καθώς είναι καταχωρημένες ήδη στον πίνακα συμβόλων .

## 2.3 Λεκτικά Κυριολεκτικά:

**Κώδικας σε Regex:**

["]([^"\\\n]|\\.|\\\n)\*["]

**Ενδεικτικά παραδείγματα από το περιβάλλον RegEx Pal:**

""

" "

"Test"

"Mark said, \"Boo!\"\n"

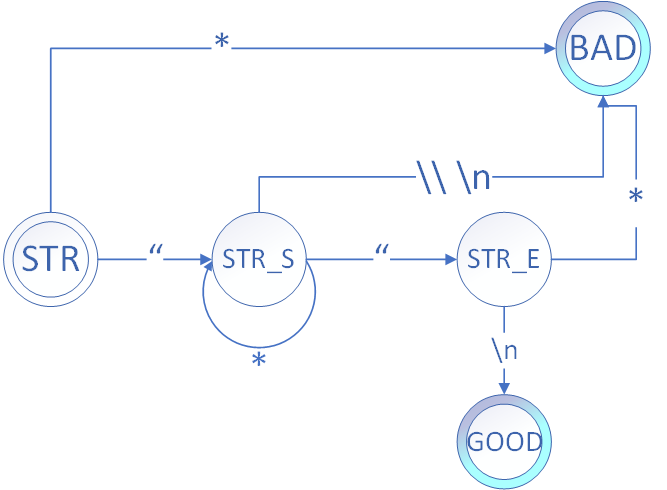
"\"

""\""

Αποδεκτά γίνονται όλα εκτός από το "\" καθώς το σκέτο backslash προκαλεί error και το ""\"" διότι παρόλο που είναι αποδεκτά τα " μέσα σε άλλα " προκαλείται error λόγω του backslash.

### 2.3.2 Πίνακας Μεταβάσεων Αυτόματου

### 2.3.1Αυτόματο



### 2.3.3Κώδικας FSM

//οταν χρησιμοποιειται το /" βγαζει bad ενώ πρέπει να δέχεται το " και να βγάζει good

START=STR

STR:

\* -> BAD

" -> STR\_S

STR\_S:

\* -> STR\_S

/ -> STR\_B

" -> STR\_E

\\\n -> BAD

STR\_E:

\* -> BAD

\n -> GOOD

STR\_B:

" -> STR\_S

\* -> STR\_S

GOOD(OK):

## 2.4 Ακέραιοι:

Επιτρεπτές τιμές ακεραίων αριθμών. Οι δεκαδικοί αριθμοί πρέπει να ξεκινάνε από μη μηδενικό αριθμό (1-9) ακολουθώντας μετά οποιοσδήποτε αριθμός ψηφίων (0-9), με εξαίρεση τον αριθμό 0. Οι δεκαεξαδικοί αριθμοί πρέπει πάντα να ξεκινάνε από 0X ή 0x και να ακολουθούνται ένα ή περισσότερα ψηφία (0-9,A-F). Οι οκταδικοί αριθμοί πρέπει πάντα να ξεκινάνε από 0 και ακολουθεί ένα ή περισσότερα στοιχεία (0-7).

**Κώδικας σε Regex:**

0[xX][0-9a-fA-F]+|[1-9][0-9]\*|0[0-7]+|0\b

**Ενδεικτικά παραδείγματα από το περιβάλλον RegEx Pal:**

3

214748

0x4F

0XFF42

063

0147

00

07

09

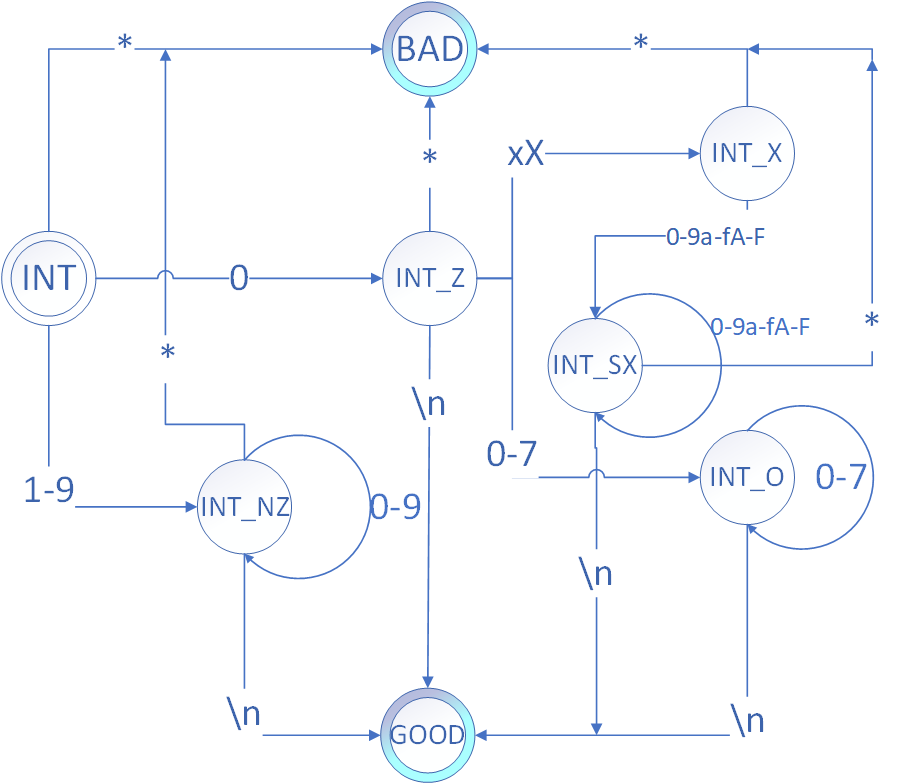
0x98xAC

Αποδεκτά γίνονται όλα εκτός από το 09 διότι το 9 δεν ανήκει στο οκταδικό οπότε θεωρείται δεκαδικός, και οι δεκαδικοί με εξαίρεση το 0 δεν επιτρέπεται να ξεκινούν με μηδενικό ψηφίο και το 0x98xAC διότι οι δεκαεξαδικοί αριθμοί ξεκινούν πάντα με 0x ή 0X όπου το x δεν μπορεί να επαναληφθεί.

### 2.4.2 Πίνακας Μεταβάσεων Αυτόματου

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Χαρακτήρες Εισόδου | | | | | | | | |
| Καταστάσεις |  | 0 | 1-7 | 8 9 | x X | a-f A-F | \* | \n |
| INT | INT\_Z | INT\_NZ | INT\_NZ |  |  |  |  |
| INT\_NZ | INT\_NZ | INT\_NZ | INT\_NZ |  |  |  | GOOD |
| INT\_Z | INT\_O | INT\_O |  | INT\_X |  |  | GOOD |
| INT\_X | INT\_SX | INT\_SX | INT\_SX |  | INT\_SX |  |  |
| INT\_SX | INT\_SX | INT\_SX | INT\_SX |  | INT\_SX |  | GOOD |
| INT\_O | INT\_O | INT\_O |  |  |  |  | GOOD |

### 2.4.1 Αυτόματο



### 2.4.3 Κώδικας FSM

START=INT

INT:1-9 -> INT\_NZ

0 -> INT\_Z

\* -> BAD

INT\_NZ:0-9 -> INT\_NZ

\n -> GOOD

\* -> BAD

INT\_Z: x X -> INT\_X

0-7 -> INT\_O

\n -> GOOD

\* -> BAD

INT\_X: 0-9 a-f A-F -> INT\_SX

\* -> BAD

INT\_SX: 0-9 a-f A-F -> INT\_SX

\n -> GOOD

\* -> BAD

INT\_O: 0-7 -> INT\_O

\n -> GOOD

\* -> BAD

GOOD(OK):

## 2.5 Αριθμοί Κινούμενης υποδιαστολής:

Επιτρεπτές τιμές αριθμών κινούμενης υποδιαστολής. Το ακέραιο μέρος του αριθμού μπορεί να είναι (0-9) όπως και το πραγματικό, χωριζόμενα με (.) . Στο πραγματικό μέρος μπορεί να εμπεριέχεται και ο αριθμός ‘e’ ή ‘E’ όπου οι αριθμοί που ακολουθούν μετά το e υψώνονται σε ακέραια δύναμη.

**Κώδικας σε Regex:**

[0-9]+((\.[0-9]+)([eE][+-]?[0-9]\*)?|([eE][+-]?[0-9]\*)+)

**Ενδεικτικά παραδείγματα από το περιβάλλον RegEx Pal:**

3.14

10.0

0.001

1e100

3.14e-10

0e0

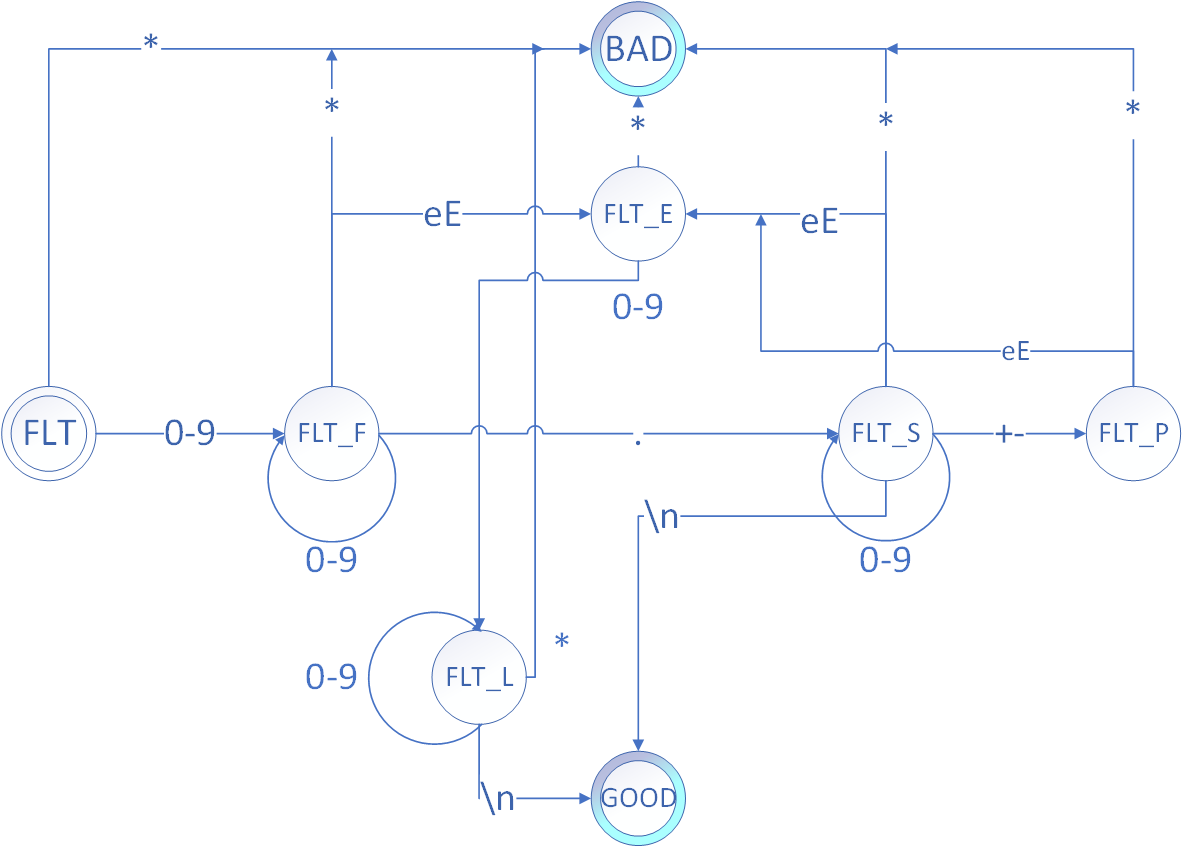
3e.6

Αποδεκτά γίνονται όλα εκτός από το 3e.6 καθώς είναι μη αποδεκτή μορφή του e.

### 2.5.2 Πίνακας Μεταβάσεων Αυτόματου

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Χαρακτήρες Εισόδου | | | | | | | |
| Καταστάσεις |  | 0-9 | e E | . | + - | \* | \n |
| FLT | FLT\_F |  |  |  |  |  |
| FLT\_F | FLT\_F | FLT\_E | FLT\_S |  |  |  |
| FLT\_S | FLT\_S | FLT\_E |  | FLT\_P |  | GOOD |
| FLT\_P |  | FLT\_E |  |  |  |  |
| FLT\_E | FLT\_L |  |  |  |  |  |
| FLT\_L | FLT\_L |  |  |  |  | GOOD |

## 2.5.1 Αυτόματο



### 2.5.3 Κώδικας FSM

START=FLT

FLT: 0-9 -> FLT\_F

\* -> BAD

FLT\_F: 0-9 -> FLT\_F

e E -> FLT\_E

. -> FLT\_S

\* -> BAD

FLT\_S: 0-9 -> FLT\_S

e E -> FLT\_E

\n -> GOOD

\* -> BAD

FLT\_P: 0-9 -> FLT\_L

\* -> BAD

FLT\_E: 0-9 -> FLT\_L

+ \- -> FLT\_P

\* -> BAD

FLT\_L: 0-9 -> FLT\_L

\n -> GOOD

\* -> BAD

GOOD(OK):

## 2.6 Τελεστές:

Σύμφωνα με την εκφώνηση για όσο αφορά τους τελεστές δεν απαιτούνται κανονικές εκφράσεις Κ.Ε. καθώς είναι καταχωρημένοι ήδη στον πίνακα συμβόλων.

## 2.7 Σχόλια:

Στη Uni-C υπάρχουν τα σχόλια γραμμής που συμβολίζονται με // και ολοκληρώνονται με το τέλος της γραμμής \n. Και τα σχόλια πολλαπλών γραμμών που συμβολίζονται με /\*(περιεχόμενο) \*/.

**Κώδικας σε Regex:**

^//.\*|\/\\*(\\*(?!\/)|[^\*])\*\\*\/

**Ενδεικτικά παραδείγματα από το περιβάλλον RegEx Pal:**

//example1/\*

/\*example2 .1

example 2.2

example 2.3\*/

/\*example3\*/

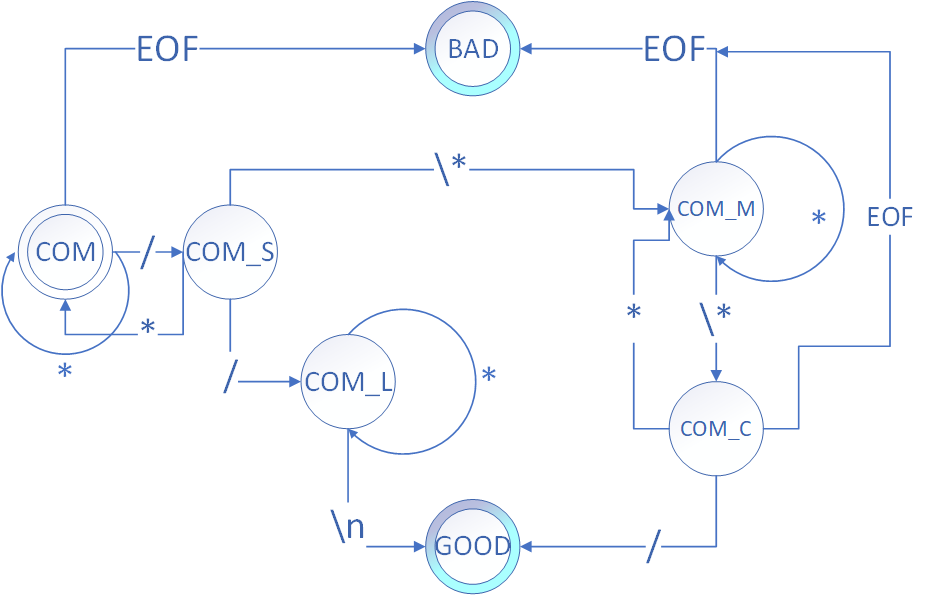
/\*example//example

Αποδεκτά γίνονται όλα εκτός από το τελευταίο καθώς έχει ξεκινήσει σχόλιο πολλαπλών γραμμών που δεν είναι ολοκληρωμένο οπότε και το // δεν αναγνωρίζεται ως σχόλιο γραμμής αλλά σαν απλά σύμβολα.

### 2.7.2 Πίνακας Μεταβάσεων Αυτόματου

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Χαρακτήρες Εισόδου | | | | | | |
| Καταστάσεις |  | / | \\* | \* | \n | EOF |
| COM | COM\_S | COM | COM | COM | BAD |
| COM\_S | COM\_L | COM\_M | COM | COM | COM |
| COM\_L | COM\_L | COM\_L | COM\_L | GOOD | GOOD |
| COM\_M | COM\_M | COM\_C | COM\_M | COM\_M | BAD |
| COM\_C | GOOD | COM\_M | COM\_M | COM\_M | BAD |

### 2.7.1 Αυτόματο



### 2.7.3 Κώδικας FSM

START = COM

COM: \* -> COM

/ -> COM\_S

EOF -> BAD

COM\_S: \* -> COM

/ -> COM\_L

\\* -> COM\_M

COM\_L: \* ->COM\_L

\n -> GOOD

COM\_M: \* -> COM\_M

EOF -> BAD

\\* -> COM\_C

COM\_C: \* -> COM\_M

/ -> GOOD

EOF -> BAD

GOOD(OK):

## 2.8 White\_spaces χαρακτήρες:

White spaces θεωρούνται τα απλά κενά τα tabs ή ο συνδυασμός και των δύο.

**Κώδικας σε Regex:**

\s

**Ενδεικτικά παραδείγματα από το περιβάλλον RegEx Pal:**

ex\_ample

# Ενιαίο Αυτόματο

### Αυτόματο

### Πίνακας Μεταβάσεων Αυτόματου

### Κώδικας FSM

# Σχόλια: