

## Εργασία Εξαμήνου

### Μέρος Α' – Δημιουργία Ανεξάρτητου Λεκτικού Αναλυτή με τη γεννήτρια FLEX

#### Περιγραφή

Για να μπορέσετε να προχωρήσετε στην ανάπτυξη της εργασίας του Μέρους Α' θα πρέπει να έχετε αποκτήσει τις απαραίτητες θεωρητικές γνώσεις για την **περιγραφή γραμματικών γλωσσών προγραμματισμού** και για τον τρόπο λειτουργίας ενός **λεκτικού αναλυτή** που έχουν διδαχθεί στο θεωρητικό μέρος του μαθήματος.

Για να βοηθηθείτε στο σχεδιασμό κανονικών εκφράσεων χρησιμοποιήστε το εργαλείο regexpal στον ιστοτόπο <https://www.regexpal.com/> ή άλλο ανάλογο που θα βρείτε μέσα στο εργαλείο [Σύνδεσμοι](#) στο eClass.

Για να βοηθηθείτε στον έλεγχο ορθότητας των αυτομάτων πεπερασμένων καταστάσεων που θα σχεδιάσετε χρησιμοποιήστε το μετα-εργαλείο fsm που είναι διαθέσιμο στο open eClass. Οδηγίες για την εγκατάσταση του εργαλείου και πληροφορίες για τον τρόπο ρύθμισης και λειτουργίας υπάρχουν στον φάκελο [O 1-KE και FSM](#) στο open eClass.

Για να βοηθηθείτε στη δημιουργία του λεκτικού αναλυτή μέσω της γεννήτριας FLEX, διαβάστε και κατανοήστε τις πληροφορίες στο σχετικό αναρτημένο αρχείο διαφανειών και στον φάκελο [2-FLEX](#) ή από άλλες πηγές στο διαδίκτυο. Σχετική παρουσίαση θα πραγματοποιηθεί στο αντίστοιχο μάθημα του εργαστηρίου.

Αφού ολοκληρώσετε την εργασία, θα πρέπει να είστε σε θέση να:

- σχεδιάζετε κανονικές εκφράσεις για την περιγραφή των λεκτικών μονάδων μιας γραμματικής
- δημιουργείτε πεπερασμένα αυτόματα από κανονικές εκφράσεις
- γράφετε κώδικα FLEX για τη δημιουργία ανεξάρτητου λεκτικού αναλυτή

Σκοπός αυτής της εργασίας είναι να αποκτήσετε σχετική εμπειρία στην περιγραφή γραμματικών μέσω κανονικών εκφράσεων και ειδικότερα να εξοικειωθείτε με την ανάπτυξη λεκτικών αναλυτών μέσω γεννήτριας ΛΑ. Για να αναπτύξετε τον λεκτικό σας αναλυτή χρησιμοποιήστε το FLEX το οποίο είναι ελεύθερα διαθέσιμο, ακολουθώντας τις αναρτημένες οδηγίες στον φάκελο [O 0-Linux & C](#). Θα δοκιμάσουμε τους κώδικες που θα καταθέσετε χρησιμοποιώντας την έκδοση του FLEX σε περιβάλλον Linux, οπότε βεβαιωθείτε ότι οι λύσεις σας λειτουργούν σωστά στην συγκεκριμένη έκδοση πριν τις υποβάλετε μέσω του εργαλείου των Εργασιών του open eClass.

Για να ολοκληρώσετε το μέρος Α' της εργασίας σας θα πρέπει να απαντήσετε στα παρακάτω επιμέρους μέρη στα χρονικά πλαίσια που έχουν τεθεί για το καθένα:

#### ΜΕΡΟΣ Α-1: Εκμάθηση σύνταξης ΚΕ

Το πρώτο μέρος αυτής της εργασίας είναι να πειραματιστείτε στο περιβάλλον regex ή άλλο ανάλογο, και να περάσετε από μια αυτό-εκπαίδευση μέσω παραδειγμάτων που θα βρείτε μέσα στο αρχείο **Εξάσκηση\_σε\_ΚΕ** μέσα στον φάκελο [1-Υλικό για εξάσκηση](#).

Το βήμα αυτό σχεδιάστηκε για δική σας εξάσκηση **χωρίς να χρειάζεται να καταθέσετε κάτι σχετικό στην εργασία σας.**

## **ΜΕΡΟΣ Α-2: Κωδικοποίηση αυτομάτων πεπερασμένων καταστάσεων μέσω FSM**

Σχεδιάστε τις ΚΕ που περιγράφουν τις λεκτικές μονάδες της γλώσσας της εργασίας σας, τα αντίστοιχα πεπερασμένα αυτόματα αναγνώρισής τους, τους αντίστοιχους πίνακες μεταβάσεων και το ενιαίο αυτόματο που προκύπτει (**παραλείψτε τις κρατημένες λέξεις για τις οποίες η γραμματική της γλώσσας που σας έχει δοθεί αναφέρει ότι δεν χρειάζονται πρότυπο αναγνώρισης γιατί βρίσκονται ήδη μέσα στον πίνακα συμβόλων (ή αντίστοιχη δική σας δομή), όπως π.χ. εντολές της γλώσσας).**

Προσομοιώστε το ενιαίο αυτόματο σε έναν γενικό Πίνακα Μεταβάσεων (ΠΜ) και στη συνέχεια, με τη βοήθεια του μετα-εργαλείου FSM κωδικοποιήστε τον γενικό ΠΜ για να ελέγξετε την ορθή αναγνώριση των λεκτικών μονάδων της γραμματικής.

**Σημείωση:** Το ενιαίο αυτόματο πρέπει να περιέχει ΜΟΝΟ επιτρεπτές εξόδους. Στον ενιαίο ΠΜ, αφήστε κενά τα κελιά που αντιστοιχούν σε ΜΗ επιτρεπτές μεταβάσεις και στο FSM κωδικοποιήστε τα ως μια κοινή μετάβαση με όνομα 'BAD'.

**Φροντίστε να αναρτήσετε την απάντησή σας,** καταθέτοντας ένα συμπιεσμένο αρχείο ZIP με όνομα που θα αρχίζει με τον κωδικό της ομάδας υποχρεωτικά και το οποίο θα περιλαμβάνει:

α) Ένα έγγραφο Word ή PDF που θα περιέχει:

- Εξώφυλλο (με ονοματεπώνυμο μελών ομάδας, αριθμό ομάδας & τμήμα) & Πίνακας Περιεχομένων
- Κανονικές εκφράσεις (μόνο επιμέρους – όχι ενιαία)
- Αυτόματα πεπερασμένων καταστάσεων ή ΔΜ (επιμέρους & ενιαίο)
- Πίνακες μεταβάσεων (επιμέρους & ενιαίος)
- Περιπτώσεις ελέγχου / εξαντλητικές δοκιμαστικές εκτελέσεις / αποτελέσματα και σχολιασμός τους
- Σχόλια πάνω στα παραπάνω υποβαλλόμενα (πχ. τρόπος λειτουργίας, προβλήματα που αντιμετωπίσατε, αν / πως τα λύσατε ή αν δεν τα λύσατε τι δοκιμάσατε να κάνετε
- Ρητή αναφορά ελλείψεων (σε σύγκριση με τα ζητούμενα) καθώς και αναφορά για την ορθή ή μη εκτέλεση του κώδικά σας
- Ανάλυση αρμοδιοτήτων/ρόλων όλων των μελών της ομάδας
- Φροντίστε να συμβουλευτείτε τις οδηγίες τεκμηρίωσης στο τέλος του παρόντος εγγράφου

β) Το αρχείο FSM του ενιαίου αυτόματου καθώς και τα επιμέρους αρχεία FSM που έχετε δημιουργήσει, επαρκώς σχολιασμένα

**Σημείωση:** Όλα τα παραπάνω συνίσταται να υλοποιηθούν το συντομότερο δυνατό αλλά θα υποβληθούν μαζί με το μέρος Α-3 της εργασίας

## **ΜΕΡΟΣ Α-3: Συμπλήρωση πρότυπου κώδικα FLEX**

Στο μέρος αυτό συνίσταται κατ'αρχάς να ολοκληρώσετε τον πρότυπο ημιτελή κώδικα simple-flex-code.l μέσα από το συμπιεσμένο αρχείο με όνομα simple-flex-code.zip που θα βρείτε μέσα στον φάκελο **Εκφωνήσεις Εργασιών** του φακέλου [4-Υλικό για Εργασίες](#). Το αρχείο αυτό δίνεται ως βάση για την ευκολότερη ανάπτυξη του δικού σας αναλυτή, η χρήση του όμως είναι προαιρετική καθώς μπορείτε να αναπτύξετε έναν ΛΑ από το μηδέν αρκεί η υλοποίησή σας να πληροί τις αναφερόμενες προδιαγραφές.

Για το simple-flex-code.l θα πρέπει να αντικαταστήσετε τα **"FILL ME"** μέσα στο αρχείο με τον κώδικα FLEX καθώς και στο αντίστοιχο token.h αρχείο header με τον κώδικα που λείπει. Αρχικά πραγματοποιήστε όλες τις απαιτούμενες αντικαταστάσεις και διορθώστε τυχόν προβλήματα που ενδεχομένως υπάρχουν στον κώδικα. Ακολουθώντας δοκιμάστε την εκτέλεσή του αρχείου σας με τα λεξήματα που δίνονται στο αρχείο input.txt, συγκρίνετε τα αποτελέσματα της εκτέλεσης με αυτά που δίνονται στο αρχείο output.txt και

προσπαθήστε να κατανοήσετε τη λειτουργία της γεννήτριας. Για την σωστή υλοποίηση του ζητούμενου ακολουθήστε τα παρακάτω βήματα:

1. Κατεβάστε το αρχείο **simple-flex-code.zip** από τις [ΕΚΦΩΝΗΣΕΙΣ ΕΡΓΑΣΙΩΝ](#) στο eClass.
2. Αποθηκεύστε / μεταφέρετε το αρχείο στο home directory σας
3. Αποσυμπιέστε το αρχείο δίνοντας την εντολή: **unzip simple-flex-code.zip**
4. Δώστε στο τερματικό την εντολή **make** προκειμένου να γίνει μεταγλώττιση (compilation) και να τρέξει αυτόματα ο λεκτικός αναλυτής με είσοδο το αρχείο **input.txt**

Αφού πετύχετε τα παραπάνω βήματα, **ολοκληρώστε τον λεκτικό αναλυτή** σας ώστε να αναγνωρίζει όλες τις λεκτικές μονάδες, τα διαχωριστικά και τα σχόλια της γλώσσας που σας έχει δοθεί. Φροντίστε επίσης απαραιτήτως να εισάγετε τα δικά σας σχόλια στον κώδικα που θα γράψετε (μέσα στο αρχείο .I) χρησιμοποιώντας τις γνώριμες από τη γλώσσα C ακολουθίες χαρακτήρων /\* και \*/ με μεγάλη όμως προσοχή ώστε να μη δημιουργήσετε πρόβλημα στη διαδικασία παραγωγής του αρχείου C από τον κώδικα FLEX.

Ο λεκτικός αναλυτής θα πρέπει επιπλέον:

- να αγνοεί τους `white_spaces` χαρακτήρες μετά το διαχωρισμό τους
- να αναγνωρίζει τα σχόλια και να τα αγνοεί
- όταν συναντάει χαρακτήρα αλλαγής γραμμής να αυξάνεται ένας μετρητής γραμμών που είναι απαραίτητος για να εμφανίζεται σε ποια γραμμή βρίσκεται η συμβολοσειρά που αναλύεται
- όταν αναγνωρίζει και προσδιορίζει επιτυχώς μια λεκτική μονάδα (token), να επιστρέφει τη γραμμή του πηγαιού κώδικα μέσα στην οποία εντοπίστηκε, το αναγνωριστικό όνομα του αντίστοιχου token και την τιμή του π.χ. `Line=1, token=INTEGER, value="5"`.
- σε οποιαδήποτε άλλη περίπτωση να σαρώνει όλους τους χαρακτήρες της άγνωστης συμβολοσειράς μέχρι να βρει διαχωριστή και να εμφανίζει το παρακάτω μήνυμα:  
`Line= αριθμός γραμμής, UNKNOWN TOKEN, value="τιμή λανθασμένης συμβολοσειράς"`  
(και να συνεχίσει κανονικά με την ανάλυση της επόμενης λέξης).

Αναρτήστε έγκαιρα τον ολοκληρωμένο πρότυπο κώδικά σας, σωστά σχολιασμένο για να γίνεται κατανοητός, μαζί με τα αποτελέσματα των δοκιμών σας. Η κατάθεσή σας θα είναι ένα συμπίεσμένο αρχείο ZIP με όνομα που θα αρχίζει με τον κωδικό της ομάδας υποχρεωτικά και που θα περιλαμβάνει τα παρακάτω:

1. Όλα τα βασικά αρχεία του λεκτικού αναλυτή (κώδικας flex, αρχείο header) επαρκώς σχολιασμένα
2. Συνοδευτικά αρχεία εισόδου - εξόδου (input.txt - output.txt) που θα περιλαμβάνουν αρκετές περιπτώσεις αναγνώρισης για όλες τις λεκτικές μονάδες καθώς και περιπτώσεις μη έγκυρων λεκτικών μονάδων
3. Ένα Makefile που θα επιτρέπει την αυτόματη μεταγλώττιση και εκτέλεση του κώδικά σας, όπως αυτό που παρέχεται στον πρότυπο κώδικα που κατεβάσατε
4. Ένα έγγραφο Word ή PDF που θα περιέχει:
  - a. Εξώφυλλο (με ονοματεπώνυμο μελών ομάδας, αριθμό ομάδας & τμήμα) & πίνακα περιεχομένων
  - b. Αντιπαραβολή της εισόδου που δοκιμάσατε με την έξοδο που λάβετε ώστε να φαίνεται ξεκάθαρα η σωστή λειτουργία του λεκτικού σας αναλυτή. Οι περιπτώσεις ελέγχου / δοκιμαστικές εκτελέσεις θα πρέπει να είναι εξαντλητικές.

- c. Αναλυτικό σχολιασμό των αποτελεσμάτων των δοκιμών (συμβουλευτείτε τις οδηγίες τεκμηρίωσης στο τέλος του παρόντος εγγράφου). Σε κάθε περίπτωση ένα απλό copy-paste των αποτελεσμάτων ΔΕΝ είναι αρκετό!
- d. Σχόλια πάνω στον υποβαλλόμενο κώδικα (πχ. τρόπος λειτουργίας του ΛΑ, προβλήματα που αντιμετωπίσατε, αν / πως τα λύσατε ή αν δεν τα λύσατε τι δοκιμάσατε να κάνετε
- e. Ρητή αναφορά ελλείψεων (σε σύγκριση με τα ζητούμενα) καθώς και αναφορά για την ορθή ή μη μεταγλώττιση και εκτέλεση του κώδικά σας
- f. Ανάλυση αρμοδιοτήτων/ρόλων όλων των μελών της ομάδας
- g. Φροντίστε να συμβουλευτείτε τις οδηγίες τεκμηρίωσης στο τέλος του παρόντος εγγράφου

**Σημείωση:** Ο παραπάνω κώδικας, θα αποτελέσει τη βάση για την ολοκληρωμένη ανάπτυξη του κώδικα FLEX της τελικής εργασίας σας και θα συμμετέχει ανάλογα στην βαθμολογία της.

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