

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 με τη σημειογραφία EBNF
- 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 2.6.4 του FLEX στο περιβάλλον Linux που είναι εγκατεστημένο στο εργαστήριο, οπότε βεβαιωθείτε ότι οι λύσεις σας λειτουργούν σωστά στο FLEX 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 στον [Προγραμματισμό Μαθημάτων](#):

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 σε EBNF τις ΚΕ που περιγράφουν τις λεκτικές μονάδες της γλώσσας της εργασίας σας, τα αντίστοιχα πεπερασμένα αυτόματα αναγνώρισής τους και το ενιαίο αυτόματο που προκύπτει από αυτά (**παραλείψτε τις κρατημένες λέξεις, 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 ελέγχετε την ορθή αναγνώριση των λεκτικών μονάδων της γραμματικής.

Σημ. 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 έγκαιρα την απάντησή σας, καταθέτοντας ένα έγγραφο Word ή PDF (όχι σε συμπιεσμένη μορφή) το οποίο θα περιέχει τις κανονικές εκφράσεις που σχεδιάσατε, το γράφημα του αντίστοιχου ενιαίου αυτόματος πεπερασμένων καταστάσεων, τον κώδικα fsm του ενιαίου αυτόματος και όλες τις απαραίτητες περιπτώσεις ελέγχου με τα αποτελέσματά τους (με copy-paste από το περιβάλλον του fsm). Στο έγγραφο αυτό θα πρέπει ακόμη να συμπεριλάβετε εξώφυλλο με τα στοιχεία σας (ονοματεπώνυμο μελών ομάδας και τμήμα) και σχόλια πάνω στα παραπάνω υποβαλλόμενα (συμβουλευτείτε τις οδηγίες τεκμηρίωσης στο τέλος του παρόντος εγγράφου)

Σημ.: όλα τα παραπάνω θα αποτελέσουν μέρος της τεκμηρίωσης της εργασίας που θα καταθέσετε στο τέλος του εξαμήνου και θα συμμετέχουν στην βαθμολογία της (βλέπε σχετικές πληροφορίες στο τέλος της εργασίας Μέρους Β').

Αν θέλετε να σας δοθούν αναλυτικά σχόλια ανάδρασης για την προσπάθειά σας, φροντίστε να παρουσιαστείτε στο 1ο (προαιρετικό) εργαστήριο υποστήριξης ομάδων για την υλοποίηση του Μέρους Α' της εργασίας εξαμήνου.

ΜΕΡΟΣ Α-3: συμπλήρωση πρότυπου κώδικα flex

Στο μέρος αυτό θα χρειαστεί 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](#). Ο πρότυπος κώδικας αφορά την αναγνώριση ακέραιων και μεταβλητών στην μορφή ακριβώς που είχε δοθεί προς επίλυση στην Εργασία μέρους Α.

Πιο συγκεκριμένα, 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. Αφού πραγματοποιήσετε όλες τις αντικαταστάσεις, δοκιμάστε την εκτέλεσή του αρχείου σας με τα συντακτικά ορθά 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. Αποσυμπίεστε το αρχείο δίνοντας την εντολή: **unzip simple-flex-code.zip**
4. Ανοίξτε με LibreOffice και διαβάστε μια πιο ευανάγνωστη έκδοση του κώδικα Flex που βρίσκεται στο αρχείο **simple-flex-code.rtf**
5. 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 .l) 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 στην συμβολοσειρά αναγνωρίζει ένα λέξημα της πηγαίας γλώσσας, να επιστρέφει τη γραμμή του πηγαίου κώδικα μέσα στην οποία εντοπίστηκε, το αναγνωριστικό όνομα του αντίστοιχου token και το λέξημα π.χ. `Line=1, token=INTCONST, value="5"`.
- σε οποιαδήποτε άλλη περίπτωση να σαρώνει όλους τους χαρακτήρες της άγνωστης συμβολοσειράς μέχρι να βρει διαχωριστή, 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)
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 αντιπαραβάλλει την είσοδο που δοκιμάσατε με την έξοδο που λάβατε ώστε να φαίνεται ξεκάθαρα η σωστή λειτουργία του λεκτικού σας αναλυτή
5. Στο αρχείο αυτό, εκτός από την ύπαρξη εξώφυλλου με τα στοιχεία σας και του τμήματός σας, θα πρέπει να φροντίσετε να σχολιάσετε επαρκώς τα αποτελέσματα των δοκιμών (συμβουλευτείτε τις οδηγίες τεκμηρίωσης στο τέλος του παρόντος εγγράφου). Σε κάθε περίπτωση ένα απλό copy-paste ΔΕΝ είναι αρκετό!

Σημ: 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 (βλέπε σχετικά στο τέλος της εργασίας Μέρος Β').

Αν θέλετε να σας δοθούν αναλυτικά σχόλια ανάδρασης για την προσπάθειά σας, φροντίστε να παρουσιαστείτε στο 2ο (προαιρετικό) εργαστήριο υποστήριξης ομάδων για την υλοποίηση του Μέρους Α' της εργασίας εξαμήνου.

Οδηγίες για την τεκμηρίωση: Συντάξτε τεύχος τεκμηρίωσης με εξώφυλλο, περιεχόμενα, εισαγωγή, σώμα τεκμηρίωσης, δοκιμαστικές εκτελέσεις ΕΣ 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.

Τελευταία ενημέρωση: 2019.3.14