

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ
ΓΛΩΣΣΕΣ ΚΑΙ ΜΕΤΑΦΡΑΣΤΕΣ – ΗΥ340**

**ΕΑΡΙΝΟ ΕΞΑΜΗΝΟ 2022
ΔΙΔΑΣΚΩΝ: ΑΝΤΩΝΙΟΣ ΣΑΒΒΙΔΗΣ**

ΒΑΣΙΚΗ ΕΡΓΑΣΙΑ

ΦΑΣΗ 1η από 5

*Ανάθεση: Πέμπτη 24 Φεβρουαρίου 2022, 13:00
Παράδοση: Παρασκευή 3 Μαρτίου 2022, 24:00 (βράδυ)*

Γενικά

Στη 1^η φάση αυτή θα κατασκευάσετε έναν Λεξικογραφικό Αναλυτή (lexical analyzer) για τα λεξικογραφικά στοιχεία της γλώσσας *alpha* (η γλώσσα της εργασίας του μαθήματος, λεπτομέρειες της οποίας θα γνωρίσουμε αργότερα στο μάθημα) χρησιμοποιώντας τη γεννήτρια λεξικογραφικών αναλυτών **Lex** ή **Flex**. Τα παράγωγα της εργασίας σας θα πρέπει να είναι τα εξής δύο:

1. Η βασική συνάρτηση λεξικογραφικής ανάλυσης που θα πρέπει να έχει ακριβώς το συγκεκριμένο function prototype: `int alpha_yylex (void* yylval)`, με `yylval` να είναι παράμετρος pointer στη μνήμη μίας μεταβλητής τύπου `alpha_token_t`, στην οποία και θα αποθηκεύετε τα χαρακτηριστικά γνωρίσματα του εκάστοτε αναγνωρισμένου token.
2. Ένα πρόγραμμα οδηγό **al** (C ή C++) στο οποίο θα δίνετε ως command-line παράμετρο το όνομα ενός text αρχείου εισόδου με λεξικογραφικά στοιχεία της γλώσσας *alpha* (δεν χρειάζεται να είναι ορθό συντακτικά πρόγραμμα, άλλωστε ακόμη δεν γνωρίζετε το συντακτικό της *alpha*). Το **al** χρησιμοποιώντας την παραπάνω συνάρτηση θα εφαρμόζει λεξικογραφική ανάλυση με έξοδο είτε σε ένα αρχείο κειμένου, το οποίο θα προσδιορίζεται από το δεύτερο προαιρετικό command line argument του προγράμματος, αλλιώς απευθείας στο standard output. Η έξοδος θα είναι της μορφής (ανά γραμμή):

```
<Αριθμός γραμμής>:
#<Αύξων αριθμός token >
"<περιεχόμενο του token σε quotes>"
Κατηγορία token με κεφαλαία
```

Κατά περίπτωση, το επιπλέον χαρακτηριστικό γνώρισμα του token (για τη φάση ένα τα γνωρίσματα θα αποθηκεύονται στη μνήμη της παραμέτρου της `alpha_yylex` τύπου `alpha_token_t` - παρακάτω δίπλα στο βελάκι αναγράφεται ο τύπος των γνωρισμάτων ανά περίπτωση). Προφανώς πολλά tokens μπορούν να εκτυπωθούν για την ίδια γραμμή, και όχι μόνο ένα όπως στο παράδειγμα.

1: #1	"if"	KEYWORD	IF	← <i>enumerated</i>
3: #2	"16"	INTCONST	16	← <i>int</i>
7: #3	"x"	IDENT	"x"	← <i>char*</i>
...				

Αντικείμενο εργασίας

Παρακάτω ακολουθούν οι κατηγορίες λεξικογραφικών στοιχείων καθώς και τα αντίστοιχα λεξικογραφικά πρότυπα, για τη γλώσσα *alpha*. Αυτό που μπορείτε απλά να παρατηρήσετε προς το παρόν είναι η ιδιαίτερη απλότητα και οικονομία των λεξικογραφικών στοιχείων της γλώσσας αυτής.

Κατηγορία λεξικογραφικού στοιχείου	Πρότυπο σε άτυπη μορφή (όλα <i>case sensitive</i>)
Λέξη κλειδί (κάθε μία ως ξεχωριστή κατηγορία)	if else while for function return break continue and not or local true false nil

Τελεστής (κάθε ένας ως ξεχωριστή κατηγορία)	= + - * / % == != ++ -- > < >= <=
---	--

Ακέραια αριθμητική σταθερά	ψηφίο, ακολουθούμενο από ψηφία
----------------------------	--------------------------------

Πραγματική αριθμητική σταθερά	ακέραια σταθερά, μετά τελεία, μετά και πάλι ακέραια σταθερά
Σταθερά ομάδα χαρακτήρων (string)	" έπειτα διάφοροι χαρακτήρες ή \" και τέλος " (εάν εμφανίζονται μέσα στο string τα escape sequences \n \t \\ τότε μετατρέπονται στον ένα χαρακτήρα που αντιστοιχούν)
Σημείο στίξης (κάθε ένα ως ξεχωριστή κατηγορία)	{ } [] () ; , : :: . ..
Αναγνωριστικό όνομα	γράμμα και έπειτα διάφορα γράμματα ή ψηφία ή υπογράμμιση (underscore)
Σχόλια	// μέχρι τέλος γραμμής /* μέχρι */ με υποστήριξη φωλιασμένων σχολίων

Ομάδες εργασίας και παράδοση

Σημειώνεται αρχικά ότι με την 1^η φάση παγιώνονται και οι ομάδες για όλες τις φάσεις της βασικής εργασίας. Η εργασία, δηλ. κάθε ξεχωριστή φάση της, μπορεί να περατωθεί από της αρχικά ορισμένη ομάδα τριών (3) το πολύ ατόμων. **Το ποσοστό συμμετοχής της 1^{ης} φάσης στην βαθμολογία της βασικής εργασίας είναι 5%.** Το άριστα για την 1^η φάση είναι το 10.

Ομάδες οι οποίες θα κατασκευάσουν επιπλέον και 2^ο λεξικογραφικό αναλυτή με το «χέρι», ακολουθώντας μία από τις μεθόδους που έχουν διδαχθεί (state encoding, state transition table, logic hard-coding) ή έστω και κάποια άλλη εφόσον την τεκμηριώσουν επαρκώς και συντόμως, θα βαθμολογηθούν με επιπλέον **bonus 2 μονάδων** (δηλ. μπορούν να λάβουν μέχρι και 12 συνολική βαθμολογία με άριστα το 10). **Στις περιπτώσεις του bonus η εξέταση είναι πιο αυστηρή.**

Λεπτομέρειες για τον ηλεκτρονικό τρόπο παράδοσης θα δοθούν στα φροντιστήρια. Η εξέταση θα οριστεί για την εβδομάδα που ακολουθεί την παράδοση της 1^{ης} φάσης.

Ποσοστά βαθμολογίας στην εργασία ανά φάση

- 1^η φάση 5%
- 2^η φάση 10%
- 3^η φάση 40%
- 4^η φάση 5%
- 5^η φάση 40%