

# C-Lite | A mini-compiler example

## A. Λεκτικές Μονάδες

Οι λεκτικές μονάδες που αποτελούν και τα τερματικά σύμβολα της γραμματικής της C-Lite περιγράφονται στη συνέχεια. Σε παρένθεση – όπου χρειάζεται – δίνονται τα αντίστοιχα συμβολικά ονόματα που εμφανίζονται στη γραμματική της C-Lite.

Μαζί με τις λεκτικές μονάδες δίνεται και η περιγραφή των σχολίων της C-Lite, τα οποία όμως δεν εμφανίζονται στη γραμματική της γλώσσας.

Σημειώνεται ότι στη γλώσσα C-Lite δεν υπάρχει διάκριση μεταξύ κεφαλαίων και πεζών αλφαβητικών χαρακτήρων, εκτός αν αυτοί αποτελούν μέρος της λέξης μιας λεκτικής μονάδας `CCONST` ή `SCONST`.

### Λέξεις-κλειδιά

Οι παρακάτω λέξεις που αποτελούν ανεξάρτητες λεκτικές μονάδες της C-Lite:

`CHAR INT FLOAT VOID IF ELSE WHILE RETURN MAIN`

Άλλες λέξεις-κλειδιά δίνονται πιο κάτω ως τελεστές.

### Αναγνωριστικά (ID)

Συμβολοσειρές που αρχίζουν με προαιρετικό χαρακτήρα `'_'`, ακολουθούμενο από αλφαβητικό χαρακτήρα, ακολουθούμενο από μηδέν ή περισσότερους αλφαριθμητικούς χαρακτήρες ή χαρακτήρες `'_'`, και δεν αποτελούν λέξεις-κλειδιά. Δεν επιτρέπονται διαδοχικοί χαρακτήρες `'_'`, εκτός από την αρχή του αναγνωριστικού.

#### Αποδεκτά παραδείγματα:

- `a100_version_2`
- `__a100_version2`
- `a100_version2_`

#### Μη αποδεκτά παραδείγματα:

Σε όλα τα μη αποδεκτά παραδείγματα που δίνονται, η συμβολοσειρά δεν αναγνωρίζεται συνολικά, είναι όμως δυνατό να αναγνωρίζονται μέρη αυτής ως ανεξάρτητες λεκτικές μονάδες.

- `a100__version2`
- `100_version_2`
- `a100_version2__`
- `a100--version-2 2`

### Απλές σταθερές

#### Μη προσημασμένες ακέραιες (ICONST):

Ο μοναδικός χαρακτήρας `'0'`, που παριστάνει τη σταθερά με τιμή 0. Επίσης, ένας ή περισσότεροι αριθμητικοί χαρακτήρες, που ο πρώτος δεν είναι ο `'0'`, οπότε η τιμή που παριστάνεται είναι ο αντίστοιχος αριθμός σε δεκαδική βάση. Δεκτοί γίνονται επίσης και οι αριθμοί οκταδικής μορφής `«00.....»`

#### Αποδεκτά παραδείγματα:

- `0`
- `180`
- `00327`

#### Μη αποδεκτά παραδείγματα:

- `0180`
- `XB7`
- `0X0`
- `0091`

Μηδέν ή περισσότεροι αριθμητικοί χαρακτήρες που ακολουθούνται από το χαρακτήρα ‘.’ και τουλάχιστον έναν αριθμητικό χαρακτήρα. Δεκτοί γίνονται επίσης και οι αριθμοί οκταδικής μορφής «00.....»

- .5
- 180.100
- 7.0
- 00126.32

- 1100
- 5.

Οποιοσδήποτε εκτυπώσιμος ASCII χαρακτήρας (κωδικοί 32-126) μεταξύ δύο εμφανίσεων του ειδικού χαρακτήρα ‘\’. Επιπλέον, ειδικοί χαρακτήρες ASCII παριστάνονται με τη βοήθεια του χαρακτήρα ‘\’. Πιο συγκεκριμένα, ο χαρακτήρας LF (Line Feed) παριστάνεται ως ‘\n’, ο χαρακτήρας FF (Form Feed) ως ‘\f’, ο χαρακτήρας HT (Horizontal Tab) ως ‘\t’, ο χαρακτήρας CR (Carriage Return) ως ‘\r’, ο χαρακτήρας BS (BackSpace) ως ‘\b’ και ο χαρακτήρας VT (Vertical Tab) ως ‘\v’.

```
➤ 'a'
➤ '$'
➤ ' '
➤ ' '
➤ '\n'
➤ '\'
```

- 'ac'
- '\p'
- '\\'

Οποιαδήποτε συμβολοσειρά μεταξύ δύο εμφανίσεων του ειδικού χαρακτήρα ‘\’, συμπεριλαμβανομένης της κενής συμβολοσειράς. Ο χαρακτήρας ‘\’ και οι πιο πάνω ειδικοί χαρακτήρες ASCII παριστάνονται σε έναν ορμαθό χαρακτήρων με τη βοήθεια του χαρακτήρα ‘\’. Με κάθε άλλη χρήση του χαρακτήρα ‘\’ παριστάνεται ο χαρακτήρας που ακολουθεί. Έτσι, ο χαρακτήρας ‘\’ καθαυτός παριστάνεται ως \\.

Ειδικά όταν ο χαρακτήρας ‘\’ βρίσκεται στο τέλος της γραμμής, ο ορμαθός συνεχίζεται στην επόμενη γραμμή, χωρίς οι χαρακτήρες ‘\’ και αλλαγής γραμμής να αποτελούν μέρος αυτού.

```
"CHARACTER +"
""
"STRINGS START AND END WITH \\""
"CHARACTER \\ AT THE END OF THE LINE \\"
"EXTENDS STRING IN THE NEXT LINE\n"
```

- Λογικό Η (OROP): | |
- Λογικό ΚΑΙ (ANDOP): & &
- Τελεστές σύγκρισης ισότητας (EQUOP): == !=
- Τελεστές σύγκρισης ανισότητας (RELOP): > >= < <=

- Προσθετικοί τελεστές (ADDOP): + -
- Πολλαπλασιαστικοί τελεστές (MULOP): \* / %
- Λογικό ΟΧΙ (NOTOP): !
- Τελεστές αυξομείωσης (INCDEC): ++ --
- Τελεστής μεγέθους τύπου (SIZEOP): sizeof

### Άλλες λεκτικές μονάδες

Άλλοι χαρακτήρες ή ακολουθίες χαρακτήρων που αποτελούν ανεξάρτητες λεκτικές μονάδες είναι:  
 '(' (LPAREN), ')' (RPAREN), ';' (SEMI), ',' (COMMA), '=' (ASSIGN), '[' (LBRACK), ']' (RBRACK), '&' (REFER), '{' (LBRACE), '}' (RBRACE) (EOF)

Σε ορισμένες περιπτώσεις κάποιες από τις παραπάνω λεκτικές μονάδες ενεργούν ως τελεστές, όπως αυτό θα φανεί στη γραμματική και σημασιολογία της C-Lite.

Η λεκτική μονάδα EOF – που ως χαρακτήρας ορίζεται με ειδικό τρόπο από το κάθε σύστημα – δεν εμφανίζεται στη γραμματική της C-Lite αλλά πρέπει να παράγεται από το λεκτικό αναλυτή με τιμή 0 για τον τερματισμό της συντακτικής ανάλυσης.

### Σχόλια

Τα σχόλια στην C-Lite είναι συμβολοσειρές που ξεκινούν με “//” και εκτείνονται μέχρι το τέλος της τρέχουσας γραμμής.