

C Language

Tokens



Saurabh Shukla (MySirG)

Agenda

- ① What are tokens?
- ② Constants
- ③ variables
- ④ keywords
- ⑤ comments

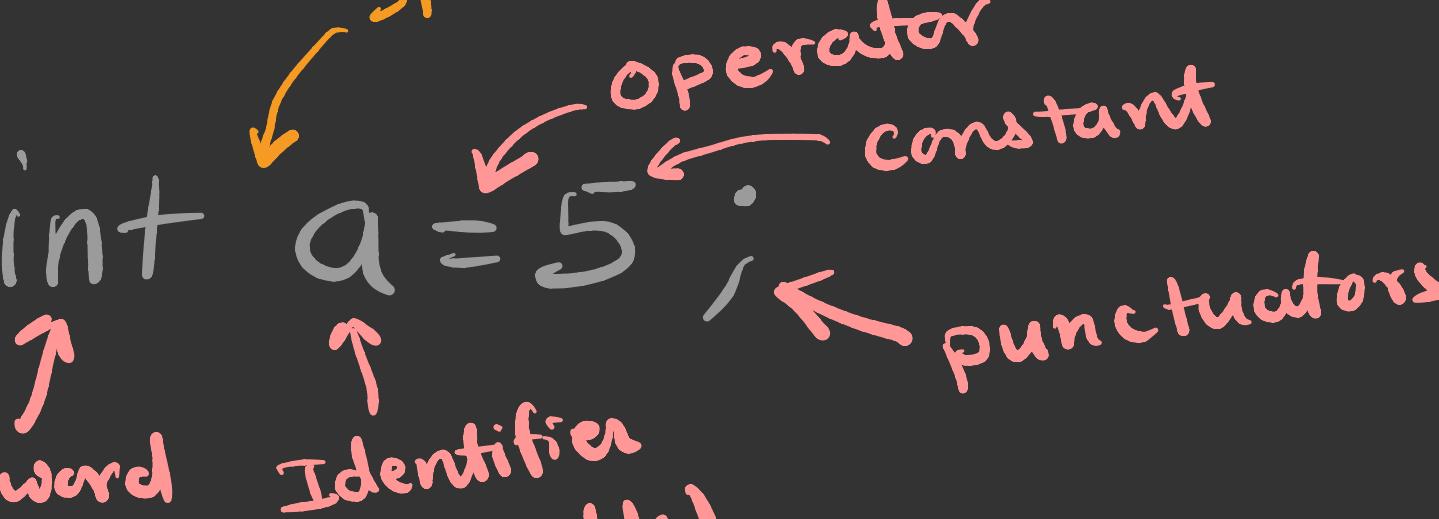
Tokens

- Tokens are also known as lexical elements
- The text of C/C++ program consists of tokens and white space.
- The token is the smallest element of a C/C++ program that is meaningful to the compiler.
- Parser is a program whose job is to separate these tokens.

→ Definition

Tokens

- Parser recognizes these kinds of tokens:
 - Constants (Literals)
 - Identifiers (variables, functions ...)
 - Keywords
 - Operators
 - Punctuators



A handwritten diagram illustrating the tokens of the C code `int a=5;`. The code is written in grey. Annotations in pink identify the tokens: "int" is labeled "Keyword", "a" is labeled "Identifier (variable)", "=" is labeled "operator", "5" is labeled "constant", ";" is labeled "punctuators", and the space between the assignment and the semicolon is labeled "space".

```
graph TD; int[int] -- Keyword --> a[a=5]; a -- Identifier --> equals[=]; equals -- operator --> five[5]; five -- constant --> semicolon[;]; semicolon -- punctuators --> space[space]
```

White Space

- Tokens are usually separated by white space , which can be one or more
 - Blanks
 - Tabs
 - New lines
 - Form feed
 - Comments

Comments

- Comments are text ignored by compiler
 - Comments are useful for documentation of your code and useful for programmers.

There are two types of comments

- ① Single line comment // text here
 - ② Multiline comment /* text line 1
text line 2
....
*/

Sample Program in C

int main()

{

 int a,b,c; //variables

 printf("Enter two numbers");

 scanf("%d %d", &a, &b);

 c=a+b;

 printf("Sum is %d", c);

}

/* main function

body ends here */

Operators

Punctuators

Constants

Identifiers

Keywords

Comments

Information = Constants = Data

primary

Integer 25 -371 0

Real 3.57 -0.06 2.0

Character 'a' 'A' '+' ' ' '1'
'3' '45' '-2'
'4.3'

Secondary

- Pointers
- array
- string "Bhopal"
- structure
- union
- enumerator

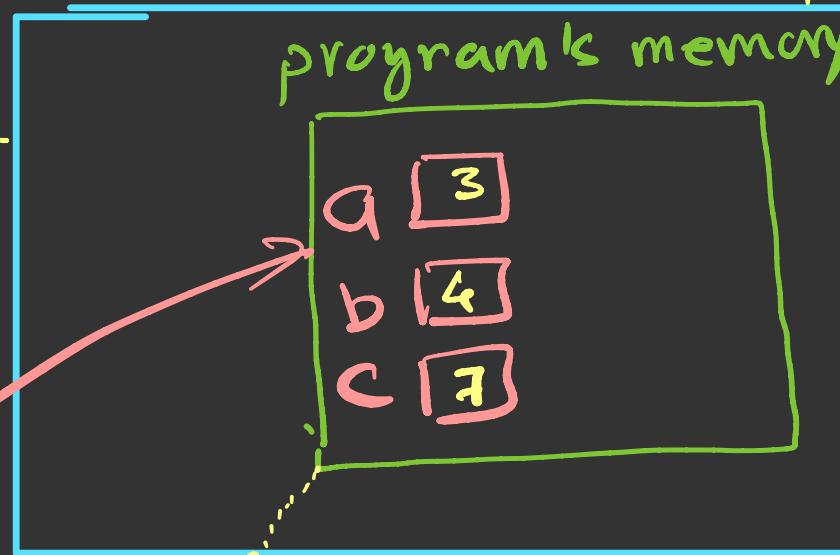
'\n' '\t' '\r'

Variables



CPU or
processor

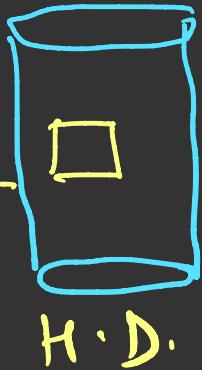
RAM



a, b, c

are variables

1 KB



Variables

- Variable is a name of memory location where we store program's data during execution of program.
- Variable name is any combination of alphabets (a to z or A to Z), digits (0 to 9) and underscore (_)
- No other symbol is allowed.
- valid variable name cannot start from a digit.
- variable name cannot be a keyword.

a_2 ✓

a_b_2_3 ✓

x-1 ✗

b\$ ✗

s.i ✗

52 ✗

x y ✗

if ✗ (keyword)

_A ✓

IF ✓

2x ✗

c language is a case sensitive
language

C is a case sensitive language

Keywords in C

predefined words / Reserved words

auto
break
case
char
const
continue
default
do

double
else
enum
extern
float
for
goto
if

int
long
register
return
short
signed
sizeof
static

struct
switch
typedef
union
unsigned
void
volatile
while

-Bool
-Complex
-Imaginary
inline
restrict

-Alignas
-Alignof
_Atomic
_Generic
_Noreturn
_Static_assert
_Thread_local

C89
32 keywords

C99
+5

C11
+7