IDENTIFIERS AND TOKENS

Character set

Like every other language, 'C' also has its own character set. A program is a set of instructions that, when executed, generate an output. The data that is processed by a program consists of various characters and symbols. The output generated is also a combination of characters and symbols.

A character set in 'C' is divided into,

- Letters
- Numbers
- Special characters
- White spaces (blank spaces)

A compiler always ignores the use of characters, but it is widely used for formatting the data. Following is the character set in 'C' programming:

1) Letters

- Uppercase characters (A-Z)
- Lowercase characters (a-z)

2) Numbers

- All the digits from 0 to 9

3) White spaces

- Blank space
- New line
- Carriage return
- Horizontal tab

4) Special characters

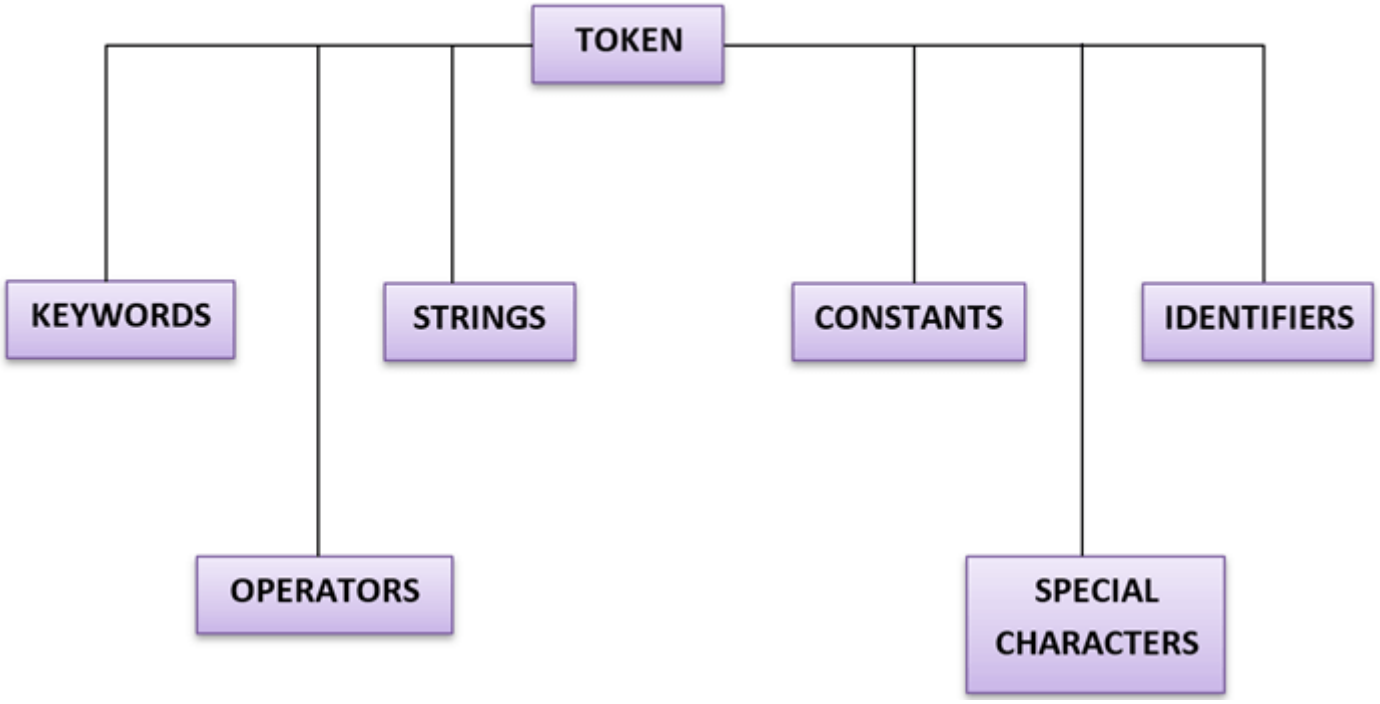
- Special characters in 'C' are shown in the given table,

Special Character	Description
, (comma)	{ (opening curly bracket)
. (period)	} (closing curly bracket)
; (semi-colon)	[(left bracket)
: (colon)	] (right bracket)
? (question mark)	((opening left parenthesis)
' (apostrophe)	) (closing right parenthesis)
" (double quotation mark)	& (ampersand)
! (exclamation mark)	^ (caret)
|(vertical bar)	+ (addition)

/ (forward slash)	– (subtraction)
\ (backward slash)	* (multiplication)
~ (tilde)	/ (division)
_ (underscore)	> (greater than or closing angle bracket)
$ (dollar sign)	< (less than or opening angle bracket)
% (percentage sign)	# (hash sign)

Token in C

TOKEN is the smallest unit in a 'C' program. It is each and every word and punctuation that you come across in your C program. The compiler breaks a program into the smallest possible units (Tokens) and proceeds to the various stages of the compilation. C Token is divided into six different types, viz, Keywords, Operators, Strings, Constants, Special Characters, and Identifiers.



Tokens in C

Keywords and Identifiers

In 'C' every word can be either a keyword or an identifier.

Keywords have fixed meanings, and the meaning cannot be changed. They act as a building block of a 'C' program. There are a total of 32 keywords in 'C'. Keywords are written in lowercase letters.

Following table represents the keywords in 'C'-

Keywords in C Programming Language			
auto	double	int	struct
break	else	long	switch
case	enum	register	typedef
char	extern	return	union
const	short	float	unsigned

continue	for	signed	void
default	goto	sizeof	volatile
do	if	static	while

An identifier is nothing but a name assigned to an element in a program. Example, name of a variable, function, etc. Identifiers in C language are the user-defined names consisting of 'C' standard character set. As the name says, identifiers are used to identify a particular element in a program. Each identifier must have a unique name. Following rules must be followed for identifiers:

1. The first character must always be an alphabet or an underscore.
2. It should be formed using only letters, numbers, or underscore.
3. A keyword cannot be used as an identifier.
4. It should not contain any whitespace character.
5. The name must be meaningful.