

Introduction of Lexical Analysis

Lexical Analysis is the first phase of the compiler also known as a scanner. It converts the High level input program into a sequence of **Tokens**.

- Lexical Analysis can be implemented with the Deterministic finite Automata.
- The output is a sequence of tokens that is sent to the parser for syntax analysis

What is a token?

A lexical token is a sequence of characters that can be treated as a unit in the grammar of the programming languages.

Example of tokens:

- Type token (id, number, real, . . .)
- Punctuation tokens (IF, void, return, . . .)
- Alphabetic tokens (keywords)

Keywords; Examples-for, while,
Identifier; Examples-Variable
Operators; Examples '+', '++',
Separators; Examples ',', ' ' 

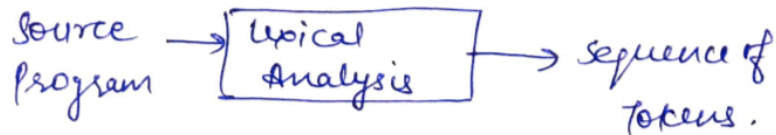
Example of Non-Tokens:

- Comments, preprocessor directive, macros, blanks, tabs, newline, etc.

Lexeme: The sequence of characters matched by a pattern to form the corresponding token or a sequence of input characters that comprises a single token is called a lexeme. eg- "float", "abs_zero_Kelvin", "=", "-", "273", ";" .

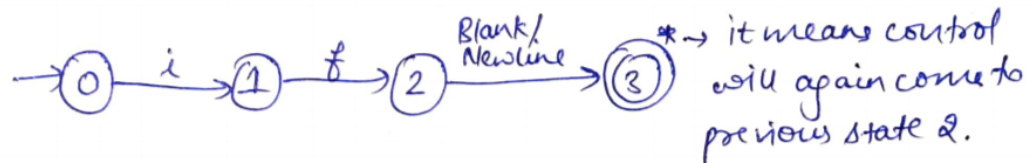
CD (Theory)

① Lexical Analysis - (Scanner)



- Lexical Analyzers can be designed using Transition Diagrams (finite Automata).

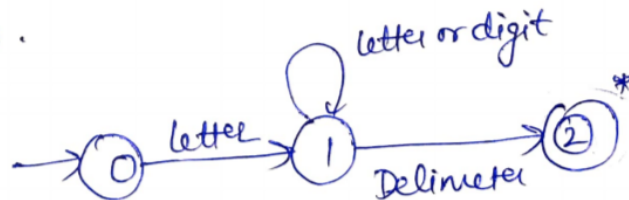
eg Transition diagram for 'if' keyword.



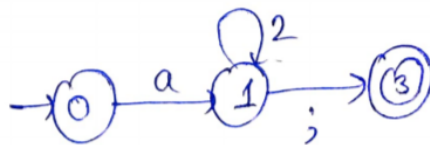
~~Transition~~

Transition diagram for an identifier -

An identifier starts with letter followed by letters or digits.



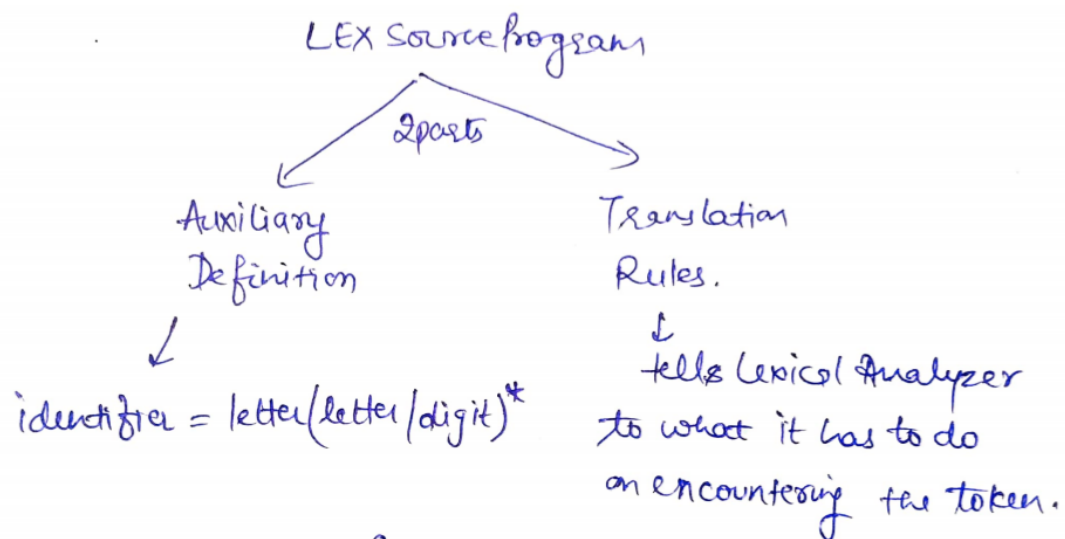
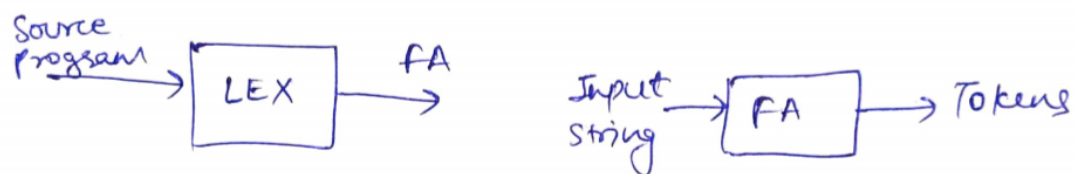
eg int a2; Here a2 is identifier



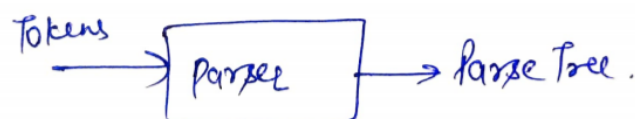
• Language for specifying Lexical Analyzer -

LEX source program represents the language used for specification of Lexical Analyzer.

LEX - It is a tool/software which automatically generates Lexical Analyzer / FA.



② Syntax Analysis (Parser) -



Lexical Analysis relates to DFA. (Later)

②

Examples -

① +++-+--44+;

No. of tokens = ?

8 tokens

② ; int a, b ; tokens ?
= 5 tokens

This is not lexical error.
It is syntax error & thus found in syntax analysis

③ for (i=0; i<=10; i++) = 13 tokens

④ int 123 ; = 3 tokens

⑤ printf ("Hello, How are you?") ; = 5 tokens

⑥ if (a>b)
{
 a=b+c;
}

= 14 tokens

⑦ if fi (a>b)
{
 a=b+c;
}

Lexical Error

⑧ fi (a,b); = 7 tokens (Here, fi is a function)
So, no error