

Practical 4

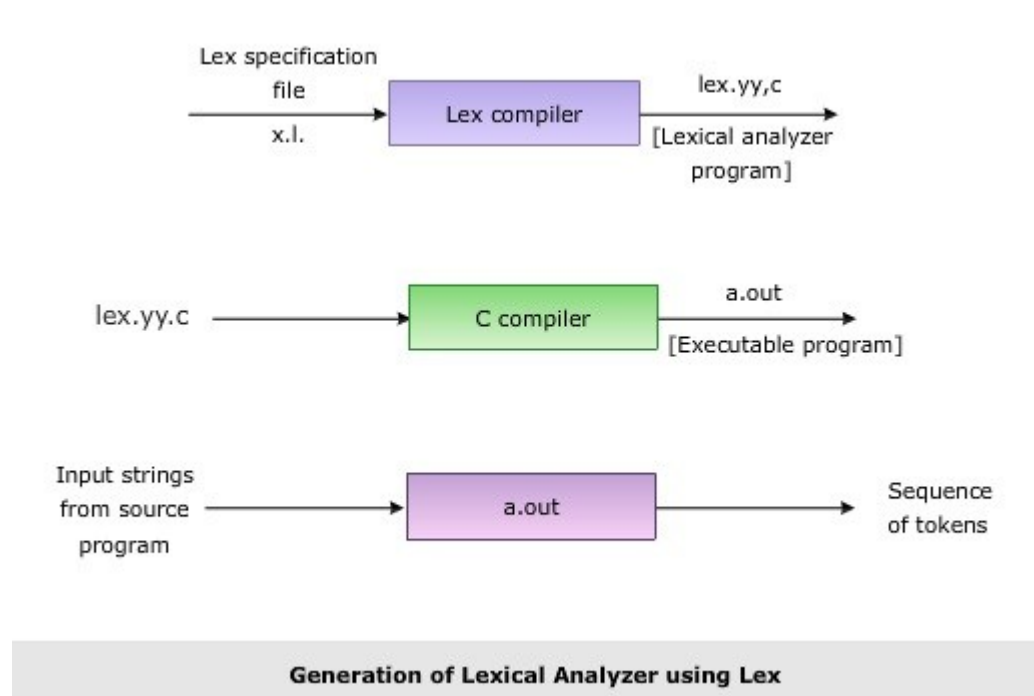
Title: To Study about Lexical Analyzer Generator (LEX) and Flex (Fast Lexical Analyzer).

➤ Lexical Analyzer Generator (LEX) Introduction

It is a tool or software which automatically generates a lexical analyzer (finite Automata). It takes as its input a LEX source program and produces lexical Analyzer as its output. Lexical Analyzer will convert the input string entered by the user into tokens as its output.

LEX is a program generator designed for lexical processing of character input/output stream. Anything from simple text search program that looks for pattern in its input-output file to a C compiler that transforms a program into optimized code.

In program with structure input-output two tasks occurs over and over. It can divide the input-output into meaningful units and then discovering the relationships among the units for C program (the units are variable names, constants, and strings). This division into units (called tokens) is known as lexical analyzer or LEXING. LEX helps by taking a set of descriptions of possible tokens n producing a routine called a lexical analyzer or LEXER or Scanner.



➤ **Lex File Format**

- A Lex program consists of three parts and is separated by % delimiters:-

Declarations

%%

Translation rules

%%

Auxiliary procedures

- **Declarations:** The declarations include declarations of variables.
- **Transition rules:** These rules consist of Pattern and Action.
- **Auxiliary procedures:** The Auxiliary section holds auxiliary functions used in the actions.

For example:

declaration

number[0-9]

%%

translation

if {return (IF);}

%%

auxiliary function

int numberSum()

➤ **Lex Predefined Function and Variables**

- **yylex()**

Purpose: This is the main function of the lexical analyzer.

Returns: An integer token to the parser (yyparse()).

Usage: Called repeatedly to tokenize input.

- **yywrap()**

Purpose: Called when end of input (EOF) is reached in yylex().

Returns: 1 if no more input, else 0.

Usage: You can use it to restart scanning or cleanup.

- **char *yytext**

Purpose: Points to the current matched text (lexeme).

Set by: Flex after each match.

Usage: Access the actual string matched.

- **yyleng**

Purpose: Contains the length of the matched string yytext.

Set by: Automatically by the scanner.

- **yyin**

Purpose: Input file pointer; tells yylex() where to read from.

Default: stdin

Set manually for custom input files.

- **yyout**

Purpose: Output file pointer; where ECHO or printed output goes.

Default: stdout

- **yyomore()**

Purpose: Tells Flex to append the next matched text to yytext.

Default behavior: Normally, new matches replace yytext.

Use when you want to build a larger token in parts.

- **yyless(n)**

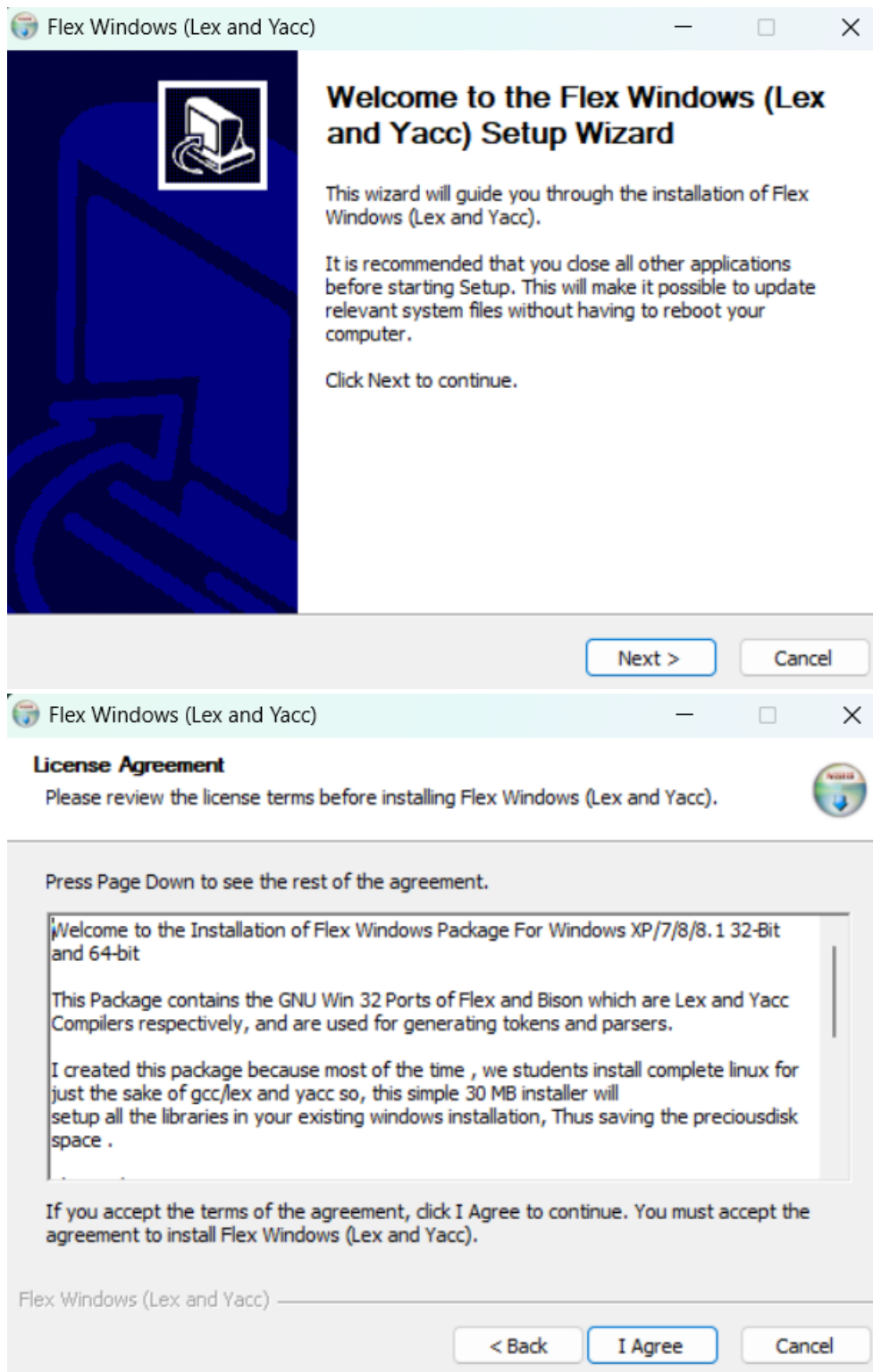
Purpose: Keeps first n characters of yytext, and pushes back the rest.

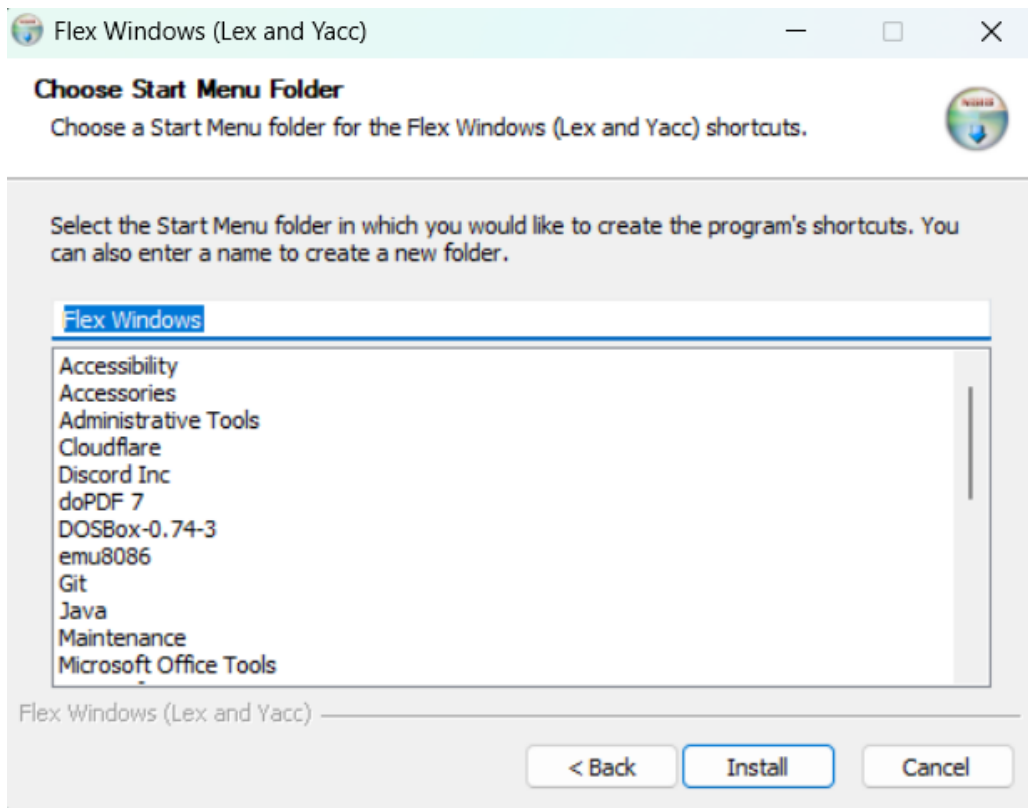
Use case: When you've matched too much and want to keep part of it.

- **yyval**

Purpose: Holds the semantic value of a token for the parser (bison/yacc).

Use: Set its value in your rules to pass data to the parser.





```
Command Prompt
Microsoft Windows [Version 10.0.26100.4652]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Lenovo>lex -V
lex version 2.5.4

C:\Users\Lenovo>
```