SNS COLLEGE OF TECHNOLOGY

Coimbatore-36.
An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A+' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

COURSE NAME : 19CSB301& AUTOMATA THEORY AND COMPILER DESIGN

III YEAR/ V SEMESTER

UNIT – II COMPILERS & LEXICAL ANALYSIS

Topic: Lex Tool Specification

Dr.B.Vinodhini
Associate Professor
Department of Computer Science and Engineering

Lex Specification

LEX

Lex is a program that generates lexical analyzer. It is used with YACC parser generator.

The lexical analyzer is a program that transforms an input stream into a sequence of tokens.

It reads the input stream and produces the source code as output through implementing the lexical analyzer in the C program.

The function of Lex is as follows:

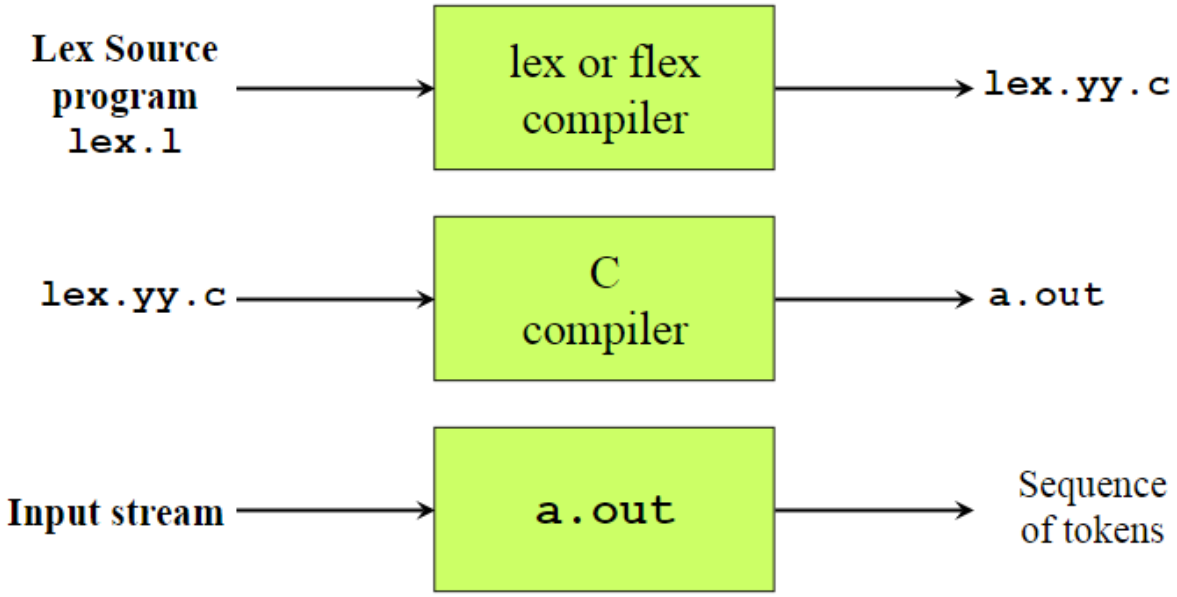
Firstly lexical analyzer creates a program lex.1 in the Lex language. Then Lex compiler runs the lex.1 program and produces a C program lex.yy.c.

Finally C compiler runs the lex.yy.c program and produces an object program a.out.

a.out is lexical analyzer that transforms an input stream into a sequence of tokens.

Lex Specification

Use of Lex



Lex Specification

Structure of Lex Programs

- A lex specification consists of three parts:

  regular definitions, C declarations in **%{ … %}**
  %%
  translation rules
  %%
  user-defined auxiliary (輔助) procedures

- The translation rules are of the form:

  p_1 $\{ action_1 \}$
  p_2 $\{ action_2 \}$
  …
  p_n $\{ action_n \}$

Lex Specification

Definitions include declarations of constant, variable and regular definitions.

Rules define the statement of form p1 {action1} p2 {action2}....pn {action}.

Where **pi** describes the regular expression and **action1** describes the actions what action the lexical analyzer should take when pattern pi matches a lexeme.

User subroutines are auxiliary procedures needed by the actions. The subroutine can be loaded with the lexical analyzer and compiled separately.

Lex Specification

Lex Predefined Variables

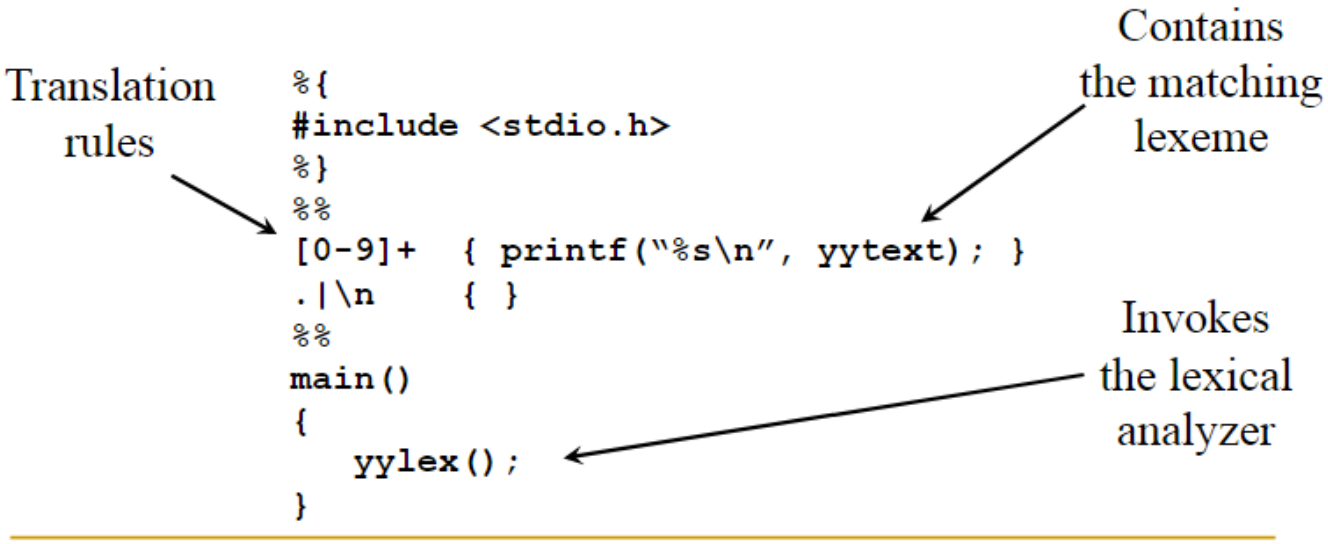
- yytext -- a string containing the lexeme
- yyleng -- the length of the lexeme
- yyin -- the input stream pointer
  - the default input of default main() is **stdin**
- yyout -- the output stream pointer
  - the default output of default main() is **stdout**.

Lex Specification

Lex Library Routines

- yylex()
  - The default main() contains a call of yylex()
- yymore()
  - return the next token
- yyless(n)
  - retain the first n characters in yytext
- yywarp()
  - is called whenever Lex reaches an end-of-file
  - The default yywarp() always returns 1

Lex Specification



Lex Specification

Regular definitions

Translation rules

```
%{
#include <stdio.h>
%}
digit      [0-9]
letter     [A-Za-z]
id         {letter}({letter}|{digit})*
%%
{digit}+   { printf("number: %s\n", yytext); }
{id}       { printf("ident: %s\n", yytext); }
.          { printf("other: %s\n", yytext); }
%%
main()
{ yylex();
}
```


Thank
you