

RAJALAKSHMI ENGINEERING COLLEGE
AN AUTONOMOUS INSTITUTION
Affiliated to ANNA UNIVERSITY
Rajalakshmi Nagar, Thandalam,
Chennai-602105



RAJALAKSHMI
ENGINEERING COLLEGE
An AUTONOMOUS Institution
Affiliated to ANNA UNIVERSITY, Chennai

DEPARTMENT OF COMPUTER SCIENCE
AND ENGINEERING

CS19641 COMPILER DESIGN LABORATORY
ACADEMIC YEAR: 2023-2024 (Even)

Name:

Roll No:

Year/ Section:

LAB PLAN

Exp No	Name of the Experiment	Time Allocation
1	Develop a lexical analyzer to recognize tokens in C. (Ex. identifiers, constants, operators, keywords etc.)	3 hrs
2	Design a Desk Calculator using LEX.	3 hrs
3	Recognize an arithmetic expression using LEX and YACC.	3 hrs
4	Evaluate expression that takes digits, *, + using YACC	3 hrs
5	Generate Three address codes for a given expression (arithmetic expression, flow of control).	3 hrs
6.a	Implement Code Optimization Techniques like dead code elimination, Common sub expression elimination	3 hrs
6.b	Implement Code Optimization Techniques like copy propagation	3 hrs
7	Generate Target Code (Assembly language) for the given set of Three Address Code.	3 hrs

Hardware requirements:PC

Software requirements:C Compiler, Editplus(FLEX)

Reference annexure 1 for installation

INDEX

S.No	Date	Experiment Name	Page No	Viva Mark

Exp No :

Date :

DEVELOP A LEXICAL ANALYZER TO RECOGNIZE TOKENS IN C

AIM:

To implement the program to identify C keywords, identifiers, operators, end statements like ; , [], { } using C tool.

ALGORITHM:

- We identify the basic tokens in c such as keywords, numbers, variables, etc.
- Declare the required header files.
- Get the input from the user as a string and it is passed to a function for processing.
- The functions are written separately for each token and the result is returned in the form of bool either true or false to the main computation function.
- Functions are issymbol() for checking basic symbols such as () etc , isoperator() to check for operators like +, -, *, / , isidentifier() to check for variables like a,b, iskeyword() to check the 32 keywords like while etc., isInteger() to check for numbers in combinations of 0-9, isnumber() to check for digits and substring().
- Declare a function detecttokens() that is used for string manipulation and iteration then the result is returned from the functions to the main. If it's an invalid identifier error must be printed.
- Declare main function get the input from the user and pass to detecttokens() function.

PROGRAM:

OUTPUT:

RESULT:

Implementation	
Output	
Viva	
Signature	

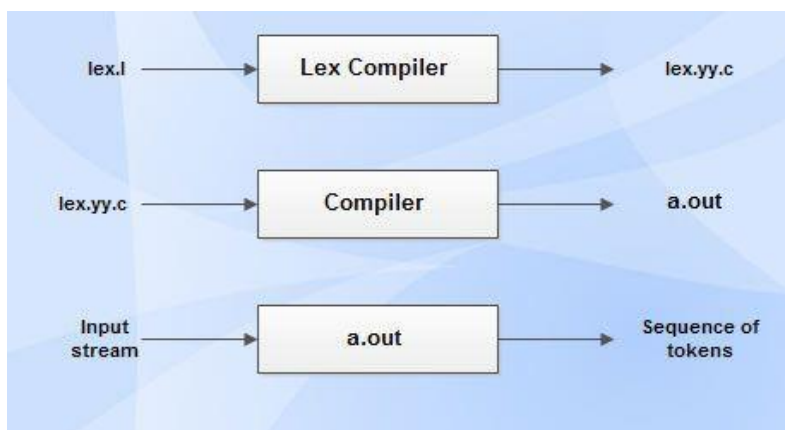
Exp No :

Date :

STUDY OF LEX TOOL

LEX:

Lex is a tool in lexical analysis phase to recognize tokens using regular expression. Lex tool itself is a lex compiler.



- lex.l is an input file written in a language which describes the generation of lexical analyzer. The lex compiler transforms lex.l to a C program known as lex.yy.c.
- lex.yy.c is compiled by the C compiler to a file called a.out.
- The output of C compiler is the working lexical analyzer which takes stream of input characters and produces a stream of tokens.
- yylval is a global variable which is shared by lexical analyzer and parser to return the name and an attribute value of token.
- The attribute value can be numeric code, pointer to symbol table or nothing.
- Another tool for lexical analyzer generation is Flex.

STRUCTURE OF LEX PROGRAMS:

Lex program will be in following form

declarations

%%

translation rules

%%

auxiliary functions

Declarations: This section includes declaration of variables, constants and regular definitions.

Translation rules: It contains regular expressions and code segments.

Form : Pattern { Action }

Pattern is a regular expression or regular definition.

Action refers to segments of code.

Patterns for tokens in the grammar

```
- digit → [0-9]
  digits → digit*
  number → digits ( . digits )? ( E [+-]? digits )?
  letter → [A-Za-z]
  id → letter ( letter | digit )*
  if → if
  then → then
  else → else
  relop → < | > | <= | >= | = | <>
- ws → (blank | tab | newline)+
```

```
%{ LT, LE, EQ, NE, GT, GE, IF,
  THEN, ELSE, ID, NUMBER, RELOP
}%
delim    [ \t\n]
ws       {delim}+
letter   [A-Za-z]
digit    [0-9]
id       {letter}({letter} | {digit})*
number   {digit}+(\.{digit}+)?(E[+-]?{digit}+)?
% %
{ws}     {}
if       { return(IF); }
then     { return(THEN); }
else     { return(ELSE); }
```

Auxiliary functions: This section holds additional functions which are used in actions. These functions are compiled separately and loaded with lexical analyzer. Lexical analyzer produced by lex starts its process by reading one character at a time until a valid match for a pattern is found. Once a match is found, the associated action takes place to produce token. The token is then given to parser for further processing.

CONFLICT RESOLUTION IN LEX:

Conflict arises when several prefixes of input matches one or more patterns. This can be resolved by the following:

- Always prefer a longer prefix than a shorter prefix.
- If two or more patterns are matched for the longest prefix, then the first pattern listed in lex program is preferred

yylex()

a function implementing the lexical analyzer and returning the token matched

yytext

a global pointer variable pointing to the lexeme matched

yylen

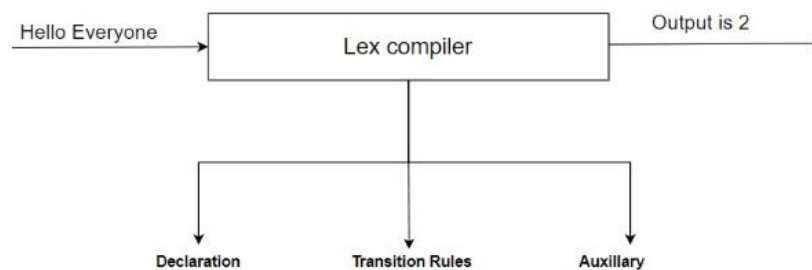
a global variable giving the length of the lexeme matched

yyval

an external global variable storing the attribute of the token

yywrap:

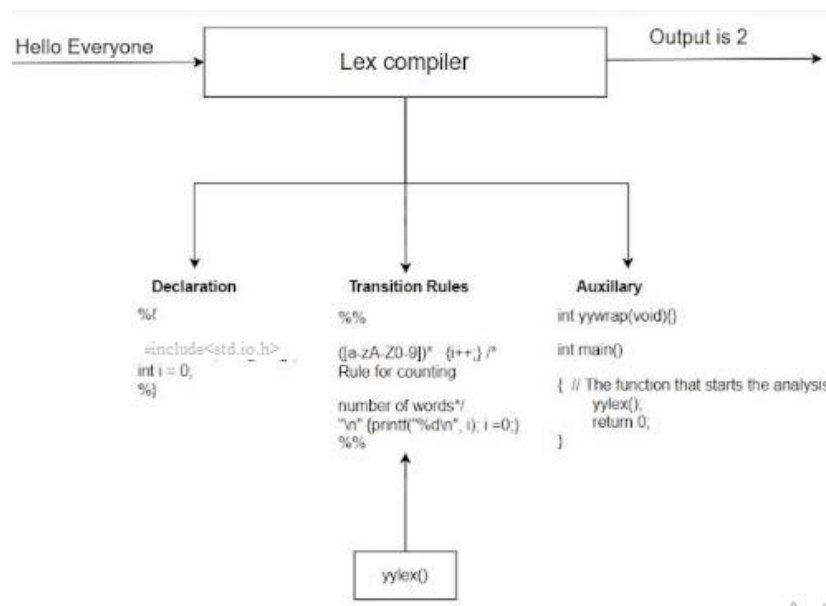
Function yywrap is called by lex when input is exhausted. Return 1 if you are done or 0 if more processing is required.



Declaration-Has header files and initialization of variables

Translation rules-write the rules for counting the words

Auxillary- has yylex() that calls the translation rules



Simple lex program:

```
/*lex program to count number of words*/

% {
#include<stdio.h>
#include<string.h>
int i = 0;
% }

/* Rules Section*/
%%
([a-zA-Z0-9])* {i++;} /* Rule for counting
number of words*/

"\n" {printf("%d\n", i); i = 0;}
%%

int yywrap(void){ }

int main()
{
    // The function that starts the analysis
    yylex();

    return 0;
}
```

```
shubham@gfg-desktop:~$ lex words.l
shubham@gfg-desktop:~$ cc lex.yy.c -lfl
shubham@gfg-desktop:~$ ./a.out
Hello Everyone
2
This is GeeksforGeeks
3
```

Implementation	
Output	
Viva	
Signature	

Exp No :

Date :

DESIGN A DESK CALCULATOR USING LEX

AIM:

Problem statement:

Create a calculator that performs addition , subtraction, multiplication and division using lex tool.

ALGORITHM:

- In the headers section declare the variables that is used in the program including header files if necessary.
- In the definitions section assign symbols to the function/computations we use along with REGEX expressions.
- In the rules section assign dig() function to the dig variable declared.
- In the definition section increment the values accordingly to the arithmetic functions respectively.
- In the user defined section convert the string into a number using atof() function.
- Define switch case for different computations.
- Define the main () and yywrap() function.

PROGRAM:

Declaration section

Rule section

RESULT:

Implementation	
Output	
Viva	
Signature	

Exp No :

Date :

STUDY OF YACC TOOL

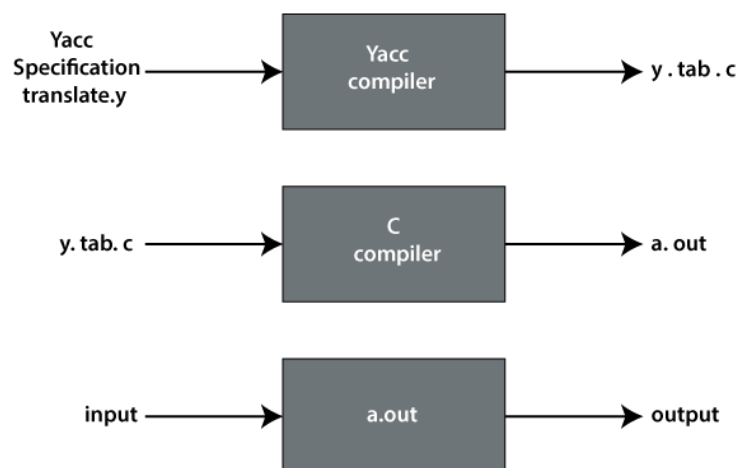
YACC TOOL:

YACC is known as Yet Another Compiler Compiler. It is used to produce the source code of the syntactic analyzer of the language produced by LALR (1) grammar. The input of YACC is the rule or grammar, and the output is a C program. If we have a file translate.y that consists of YACC specification, then the UNIX system command is:

YACC translate.y

This command converts the file translate.y into a C file y.tab.c. It represents an LALR parser prepared in C with some other user's prepared C routines. By compiling y.tab.c along with the ly library, we will get the desired object program a.out that performs the operation defined by the original YACC program.

The construction of translation using YACC is illustrated in the figure below:



STRUCTURE OF YACC:

Declarations

%%

Translation rules

%%

Supporting C rules

C declarations	<code>%{ #include <stdio.h> %}</code>
yacc declarations	<code>%token NAME NUMBER</code>
Grammar rules	<code>statement: NAME '=' expression expression { printf("= %d\n", \$1); } ; expression: expression '+' NUMBER { \$\$ = \$1 + \$3; } expression '-' NUMBER { \$\$ = \$1 - \$3; } NUMBER { \$\$ = \$1; } ; %%</code>
Additional C code	<code>int yyerror(char *s) { fprintf(stderr, "%s\n", s); return 0; } int main(void) { yyparse(); return 0; }</code>

Declarations Part: This part of YACC has two sections; both are optional. The first section has ordinary C declarations, which is delimited by `%{` and `%}`. Any temporary variable used by the second and third sections will be kept in this part. Declaration of grammar tokens also comes in the declaration part. This part defined the tokens that can be used in the later parts of a YACC specification.

```
%{
#include <stdio.h>
#include <stdlib.h>
%}
%token ID NUM
%start expr
```

Terminal

Start Symbol

Translation Rule Part: After the first `%%` pair in the YACC specification part, we place the translation rules. Every rule has a grammar production and the associated semantic action. A set of productions:

$$\langle \text{head} \rangle \Rightarrow \langle \text{body} \rangle_1 \mid \langle \text{body} \rangle_2 \mid \dots \mid \langle \text{body} \rangle_n$$

would be written in YACC as

```
<head> : <body>1    {<semantic action>1}
       | <body>2    {<semantic action>2}
       | .....
       | <body>n    {<semantic action>n}
       ;
```

Grammar rule section

```

expr  : expr '+' term
      | term
      ;
term   : term '*' factor
      | factor
      ;
factor : '(' expr ')'
      | ID
      | NUM

```

In a YACC production, an unquoted string of letters and digits that are not considered tokens is treated as non-terminals.

The semantic action of YACC is a set of C statements. In a semantic action, the symbol \$\$ is considered to be an attribute value associated with the head's non-terminal. While \$i is considered as the value associated with the grammar production of the body. If we have left only with associated production, the semantic action will be performed. The value of \$\$ is computed in terms of \$i's by semantic action.

Supporting C-Rules: It is the last part of the YACC specification and should provide a lexical analyzer named **yylex()**. These produced tokens have the token's name and are associated with its attribute value. Whenever any token like DIGIT is returned, the returned token name should have been declared in the first part of the YACC specification.

The attribute value which is associated with a token will communicate to the parser through a variable called **yylval**. This variable is defined by a YACC.

Whenever YACC reports that there is a conflict in parsing-action, we should have to create and consult the file **y.output** to see why this conflict in the parsing-action has arisen and to see whether the conflict has been resolved smoothly or not.

Instructed YACC can reduce all parsing action conflict with the help of two rules that are mentioned below:

- A reduce/reduce conflict can be removed by choosing the production which has conflict mentioned earlier in the YACC specification.
- A shift/reduce conflict is reduced in favor of shift. A shift/reduce conflict that arises from the dangling-else ambiguity can be solved correctly using this rule.

Implementation	
Output	
Viva	
Signature	

Exp No :

Date :

Recognize an arithmetic expression using LEX and YACC

AIM:

To check whether the arithmetic expression using lex and yacc.

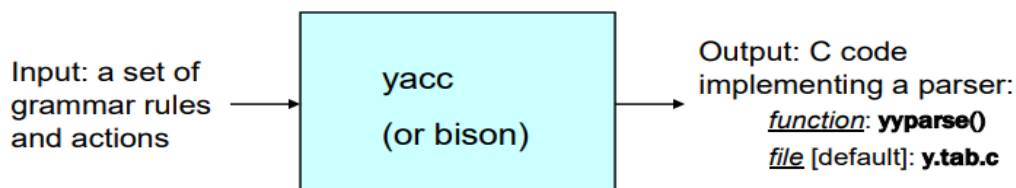
ALGORITHM:

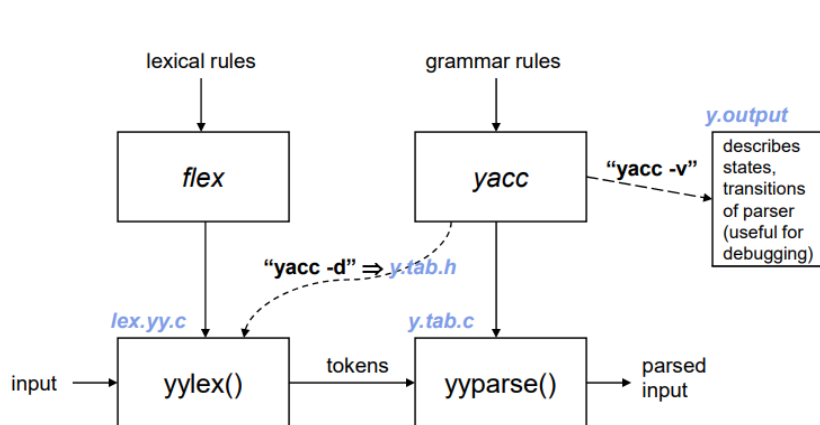
- Using the flex tool, create lex and yacc files.
- In the C include section define the header files required.
- In the rules section define the REGEX expressions along with proper definitions.
- In the user defined section define yywrap() function.
- Declare the yacc file inside it in the C definitions section declare the header files required along with an integer variable valid with value assigned as 1.
- In the Yacc declarations declare the format token num id op.
- In the grammar rules section if the starting string is followed by assigning operator or identifier or number or operator followed by a number or open parenthesis followed by an identifier. The x could be an operator followed by an identifier or operator or no operator then declare that as valid expressions by making the valid stay in 1 itself.
- In the user definition section if the valid is 0 print as Invalid expression in yyerror() and define the main function.

LEX AND YACC WORKING

Parser generator:

- Takes a specification for a context-free grammar.
- Produces code for a parser.





Yacc determines integer representations for tokens:

- Communicated to scanner in file y.tab.h
- use “yacc -d” to produce y.tab.h

Token encodings:

- “end of file” represented by ‘0’;
- A character literal: its ASCII value
- Other tokens: assigned numbers ≥ 257 .
- Parser assumes the existence of a function ‘int yylex()’ that implements the scanner.
- Scanner:
 - Return integer value indicates the type of token found
 - Values communicated to the parser using yytext, yylval
 - yytext determines lexeme of a token and yylval determines a integer assigned to a token
 - The token error is reserved for error handling

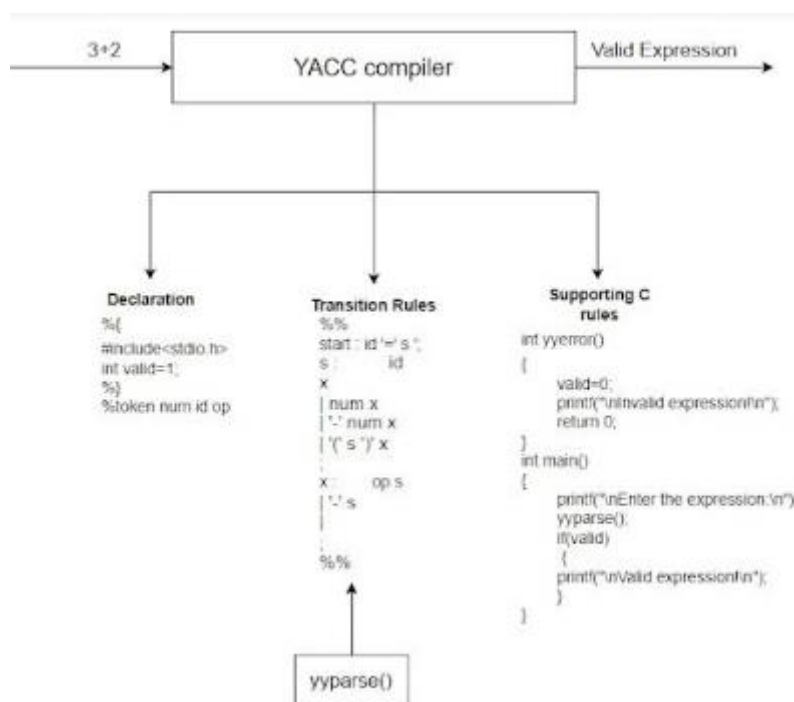
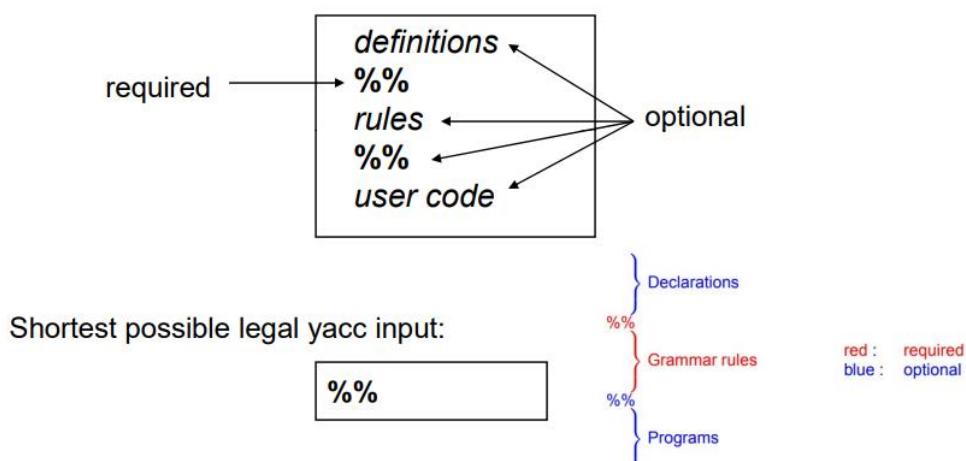
Communication between Scanner and Parser :-

Suppose the grammar specification is in the file yacc.y then

- The command ‘yacc yacc.y’ would yield a file y.tab.c containing the parser constructed by yacc
- The command ‘yacc -d yacc.y’ constructs a file y.tab.h which is included in the declaration scanner section of lex file

The user needs to supply the main() function to the driver and a file yyerror() is generated by the parser if there is an error in the input.

Yacc Input File :-



int yyparse()

Called once from main() [user-supplied]

Repeatedly calls yylex() until done:

- On syntax error, calls yyerror() [user-supplied]
- Returns 0 if all of the input was processed
- Returns 1 if aborting due to syntax error.

Example: int main() { return yyparse(); }

Yacc Grammar Rules :

Information about tokens:

- Token names:
 - Declared using ‘%token’ %token name1 name2 ...
 - Any name not declared as a token is assumed to be a nonterminal.
- Start symbol of grammar, using ‘%start’ [optional] %start name
 - If not declared explicitly, defaults to the nonterminal on the LHS of the first grammar rule listed
- Stuff to be copied verbatim into the output (e.g., declarations, #includes): enclosed in % { ... } %

Rules:

Grammar Production	Yacc Rule
A -> B1 B2...Bm	A: B1 B2 ... Bm;
B -> C1 C2,...Cn	B: C1 C2 ... Cn;
C -> D1 D2,...Dk	C: D1 D2 ... Dk;

- Rule RHS can have arbitrary C code embedded within { ... }.
E.g. A : B1 { printf(“after B1\n”); x = 0; } B2 { x++; } B3;
- Left-recursion more efficient than right-recursion:–
A : A x | ... rather than A : x A | ...

Specifying Operator Properties

Binary operators: %left, %right, %nonassoc:

Associativity of operator

%left '+' '-'

%left '*' '/'

%right '^'

Operators in the same group have the same precedence

Unary operators: %prec

- Changes the precedence of a rule to be that of the token specified. E.g.:
%left '+' '-'
%left '*' '/'
Expr: expr '+' expr

| '-' expr %prec '*'

| ...

PROGRAM:

Lex file:

Yacc file:

Definition section

Rule Section

OUTPUT:

RESULT:

Implementation	
Output	
Viva	
Signature	

Exp No :

Date :

EVALUATE EXPRESSION THAT TAKES DIGITS, *, + USING YACC

AIM:

To perform arithmetic operations using lex and yacc.

ALGORITHM:

- Using the flex tool, create lex and yacc files.
- In the definition section of the lex file, declare the required header files along with an external integer variable yylval.
- In the rule section, if the regex pertains to digit convert it into integer and store yylval. Return the number.
- In the user definition section, define the function yywrap()
- In the definition section of the yacc file, declare the required header files along with the flag variables set to zero. Then define a token as number along with left as '+', '-', 'or', '*', '/', '%' or '(')'
- In the rules section, create an arithmetic expression as E. Print the result and return zero.
- Define the following:
 - E: E '+' E (add)
 - E: E '-' E (sub)
 - E: E '*' E (mul)
 - E: E '/' E (div)
 - If it is a single number return the number.
- In driver code, get the input through yyparse(); which is also called as main function.
- Declare yyerror() to handle invalid expressions and exceptions.
- Build lex and yacc files and compile.

PROGRAM:

OUTPUT:

RESULT:

Implementation	
Output	
Viva	
Signature	

STUDY :**CODE OPTIMIZATION****CODE OPTIMIZATION:**

The process of code optimization involves

- Eliminating the unwanted code lines
- Rearranging the statements of the code

CODE OPTIMIZATION TECHNIQUES:**1. Compile Time Evaluation**

Two techniques that falls under compile time evaluation are-

A) Constant Folding

In this technique,

- As the name suggests, it involves folding the constants.
- The expressions that contain the operands having constant values at compile time are evaluated.
- Those expressions are then replaced with their respective results.

Example:

$$\text{Circumference of Circle} = (22/7) \times \text{Diameter}$$

Here,

- This technique evaluates the expression $22/7$ at compile time.
- The expression is then replaced with its result 3.14.
- This saves the time at run time.

B) Constant Propagation

In this technique,

- If some variable has been assigned some constant value, then it replaces that variable with its constant value in the further program during compilation.
- The condition is that the value of variable must not get alter in between.

Example:

$$\text{pi} = 3.14$$

$$\text{radius} = 10$$

$$\text{Area of circle} = \text{pi} \times \text{radius} \times \text{radius}$$

Here,

- This technique substitutes the value of variables 'pi' and 'radius' at compile time.
- It then evaluates the expression $3.14 \times 10 \times 10$.

- The expression is then replaced with its result 314.
- This saves the time at run time.

2. Common Sub-Expression Elimination

The expression that has been already computed before and appears again in the code for computation is called as **Common Sub-Expression**.

In this technique,

- As the name suggests, it involves eliminating the common sub expressions.
- The redundant expressions are eliminated to avoid their re-computation.
- The already computed result is used in the further program when required.

Example:

Code Before Optimization	Code After Optimization
<pre>S1 = 4 x i S2 = a[S1] S3 = 4 x j S4 = 4 x i // Redundant Expression S5 = n S6 = b[S4] + S5</pre>	<pre>S1 = 4 x i S2 = a[S1] S3 = 4 x j S5 = n S6 = b[S1] + S5</pre>

3. Code Movement

In this technique,

- As the name suggests, it involves movement of the code.
- The code present inside the loop is moved out if it does not matter whether it is present inside or outside.
- Such a code unnecessarily gets execute again and again with each iteration of the loop.
- This leads to the wastage of time at run time.

Example:

Code Before Optimization	Code After Optimization
<pre>for (int j = 0 ; j < n ; j ++) { x = y + z ; a[j] = 6 x j ; }</pre>	<pre>x = y + z ; for (int j = 0 ; j < n ; j ++) { a[j] = 6 x j ; }</pre>

4. Dead Code Elimination

In this technique,

- As the name suggests, it involves eliminating the dead code.
- The statements of the code which either never executes or are unreachable or their output is never used are eliminated.

Example:

Code Before Optimization	Code After Optimization
<pre>i = 0 ; if (i == 1) { a = x + 5 ; }</pre>	<pre>i = 0 ;</pre>

5. Strength Reduction

In this technique,

- As the name suggests, it involves reducing the strength of expressions.
- This technique replaces the expensive and costly operators with the simple and cheaper ones.

Example:

Code Before Optimization	Code After Optimization
<pre>B = A x 2</pre>	<pre>B = A + A</pre>

Here,

- The expression “A x 2” is replaced with the expression “A + A”.
- This is because the cost of multiplication operator is higher than that of addition operator.

Exp No :

Date :

**GENERATE THREE ADDRESS CODES FOR A GIVEN EXPRESSION
(ARITHMETIC EXPRESSION, FLOW OF CONTROL)**

AIM:

To generate the three address code using C program.

ALGORITHM:

- The expression is read from the file using a file pointer
- Each string is read and the total no. of strings in the file is calculated.
- Each string is compared with an operator; if any operator is seen then the previous string and next string are concatenated and stored in a first temporary value and the three address code expression is printed
- Suppose if another operand is seen then the first temporary value is concatenated to the next string using the operator and the expression is printed.
- The final temporary value is replaced to the left operand value.

PROGRAM:

RESULT:

Implementation	
Output	
Viva	
Signature	

Exp No :

Date :

IMPLEMENT CODE OPTIMIZATION TECHNIQUES LIKE DEAD CODE AND COMMON EXPRESSION ELIMINATION

AIM:

To write a C program to implement the dead code elimination and common expression elimination (code optimization) techniques.

ALGORITHM:

- Start
- Create the input file which contains three address code.
- Open the file in read mode.
- If the file pointer returns NULL, exit the program else go to 5.
- Scan the input symbol from left to right.
- Store the first expression in a string.
- Compare the string with the other expressions in the file.
- If there is a match, remove the expression from the input file.
- Perform these steps 5-8 for all the input symbols in the file.
- Scan the input symbol from the file from left to right.
- Get the operand before the operator from the three address code.
- Check whether the operand is used in any other expression in the three address code.
- If the operand is not used, then eliminate the complete expression from the three address code else go to 14.
- Perform steps 11 to 13 for all the operands in the three address code till end of the file is reached.
- Stop.

PROGRAM:

OUTPUT:

RESULT:

Implementation	
Output	
Viva	
Signature	

Exp No :

Date :

IMPLEMENT CODE OPTIMIZATION TECHNIQUES – COPY PROPAGATION

AIM:

To write a C program to implement copy propagation.

ALGORITHM:

- The desired header files are declared.
- The two file pointers are initialized one for reading the C program from the file and one for writing the converted program with constant folding.
- The file is read and checked if there are any digits or operands present.
- If there is, then the evaluations are to be computed in switch case and stored.
- Copy the stored data to another file.
- Print the copied data file.

PROGRAM:

OUTPUT:

RESULT:

Implementation	
Output	
Viva	
Signature	

Exp No :

Date :

Generate Target Code (Assembly language) for the given set of Three Address Code

AIM:

To generate assembly language for the three address code using C program.

ALGORITHM:

- Get address code sequence.
- Determine current location of 3 using address (for 1st operand).
- If the current location does not already exist, generate move (B, O).
- Update address of A (for 2nd operand).
- If the current value of B and () is null, exist.
- If they generate operator () A, 3 ADPR.
- Store the move instruction in memory.

PROGRAM:

OUTPUT:

RESULT:

Implementation	
Output	
Viva	
Signature	

Annexure 1

Download editplus from the following link:

https://drive.google.com/file/d/0B9D4jOdpRzZHNTVraV9rX280R0E/view?resourcekey=0-wKxkIdA7Eqh0j2Jj_u87ow

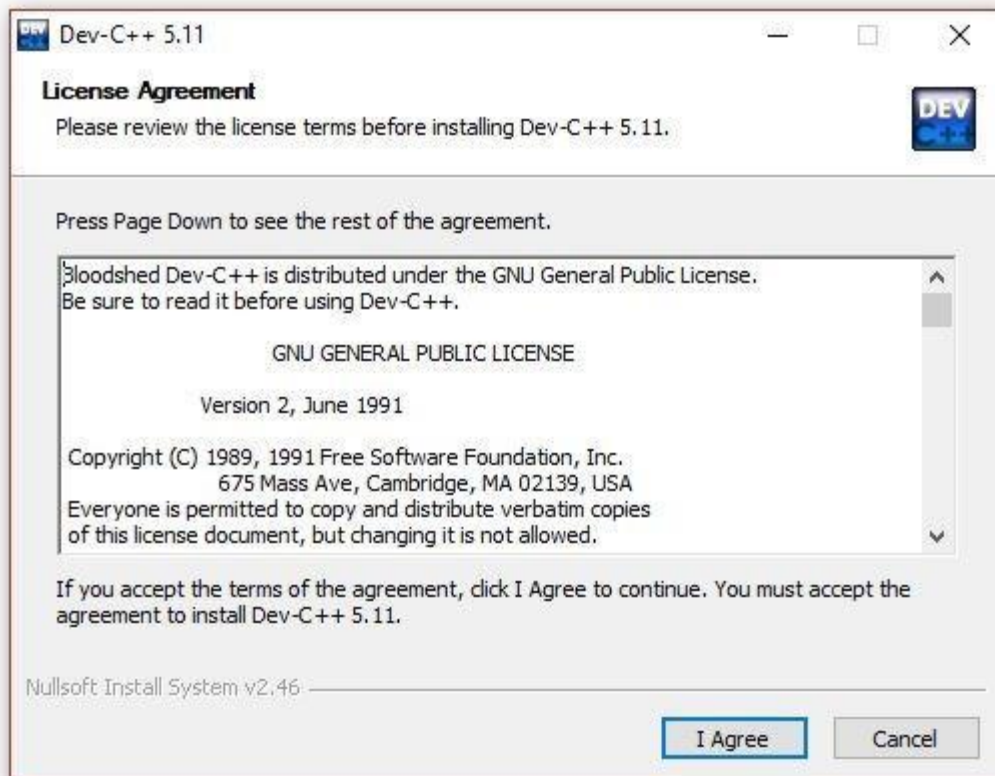
Note: Make sure that C compiler is installed.

Install the setup:

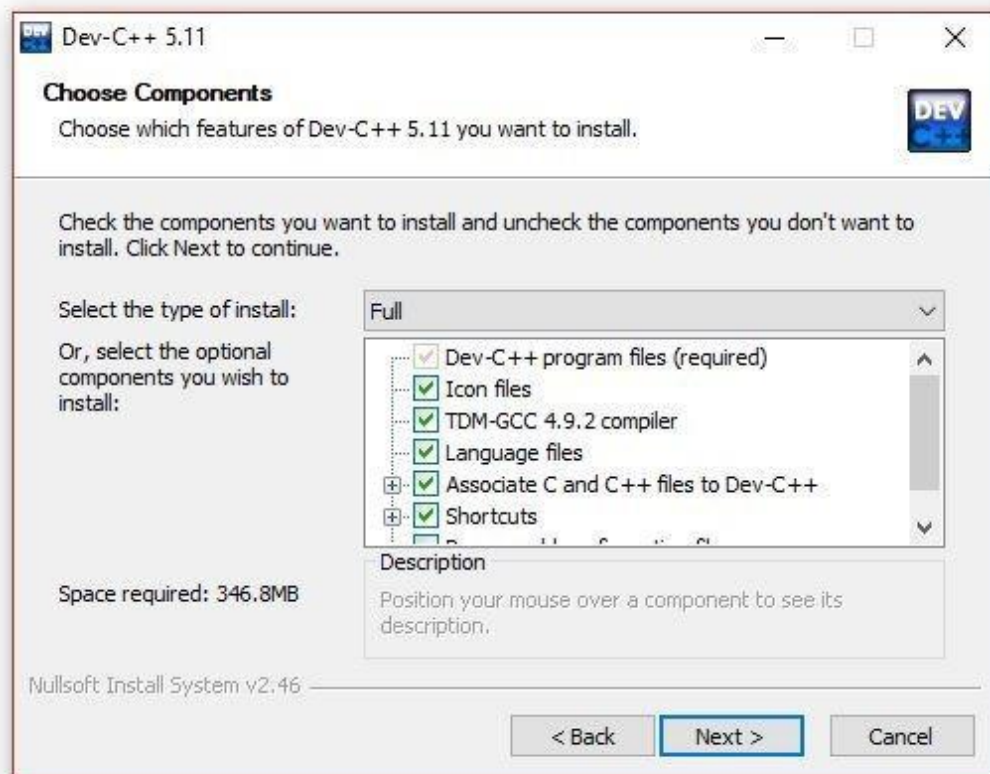
After downloading the setup double click on it, it extracts the package from the setup and asks to select the language. So select the language as per your choice and press the ok button.



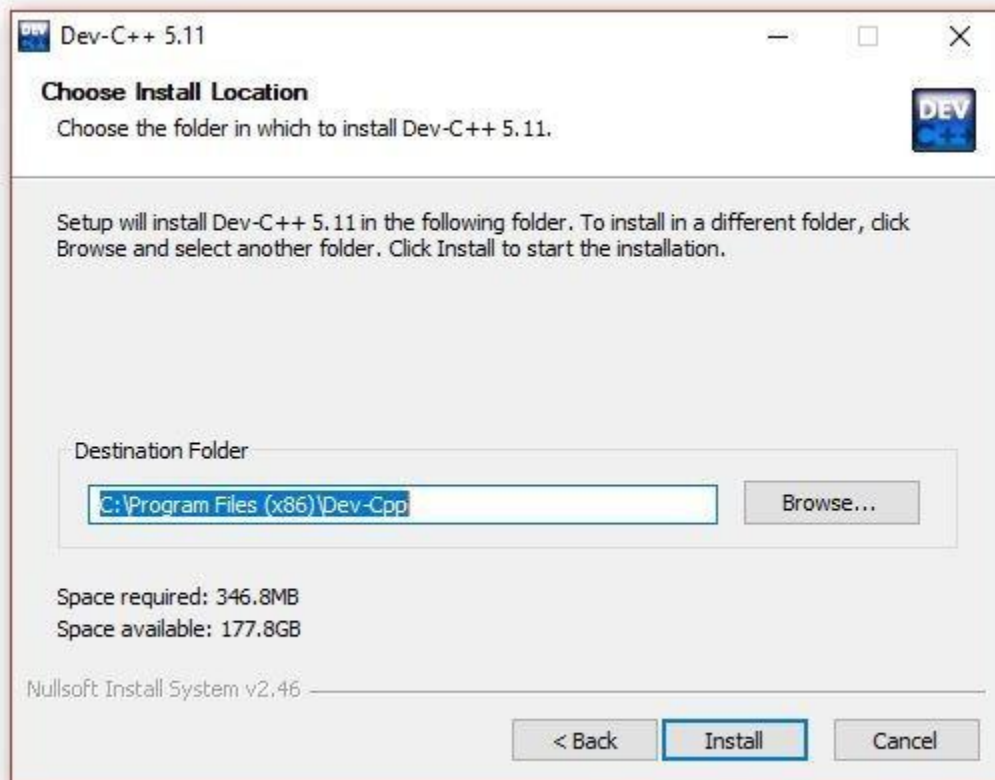
Accept the license agreement.

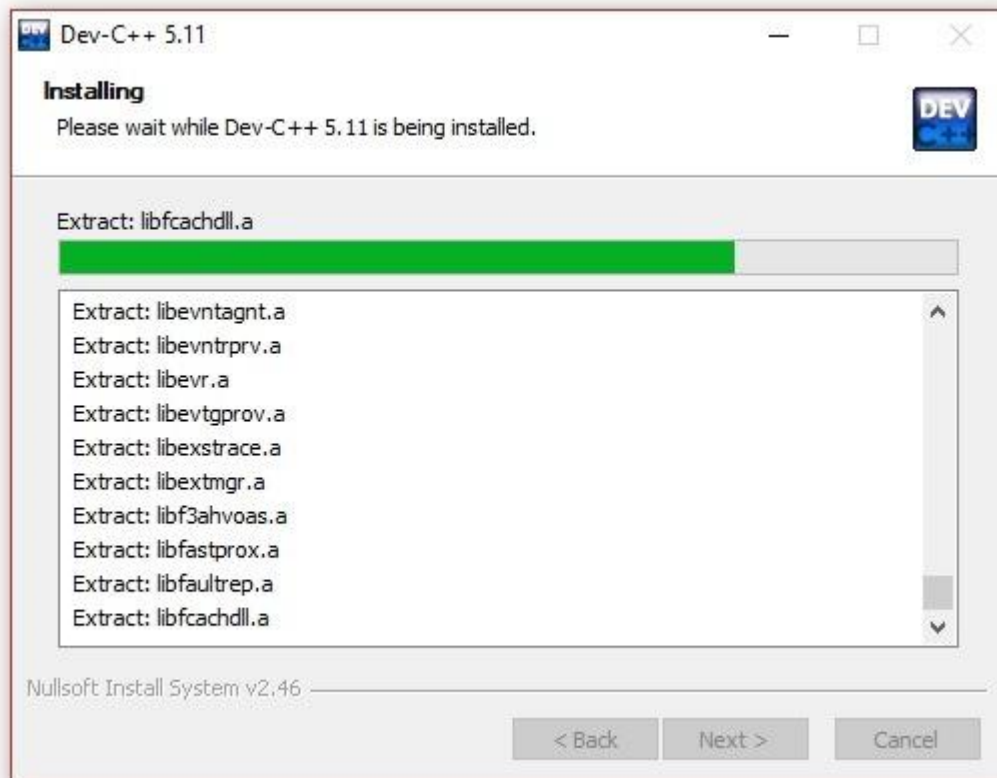


In the next prompt window, it will ask you to select the package you want to install. Here you don't need to anything just press the Next button.



Select the destination folder (keep it to default) and press the install button. It will take a few minutes.

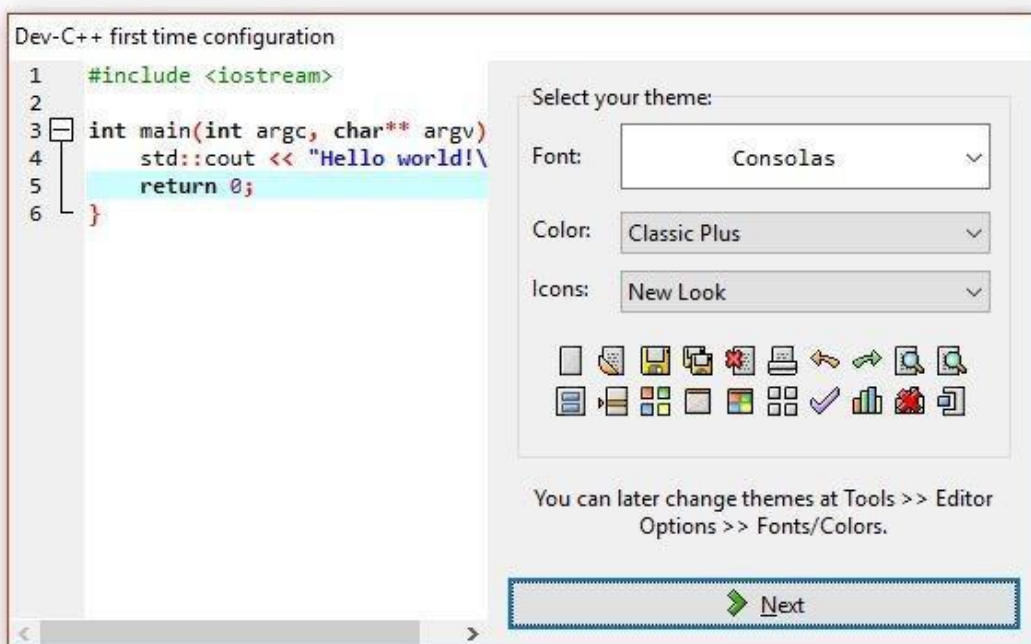
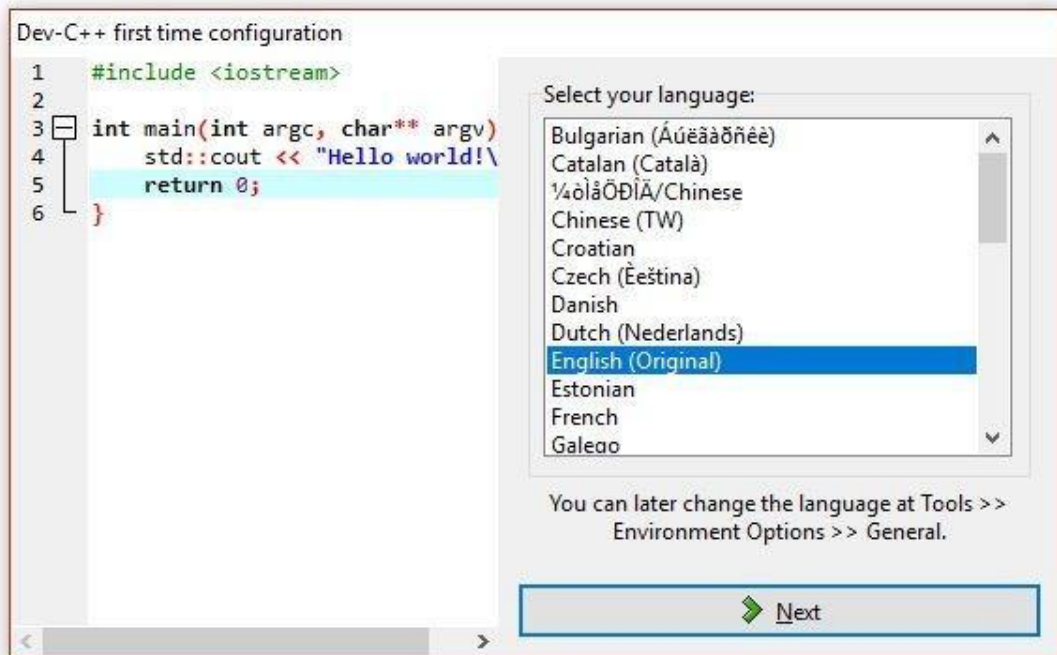


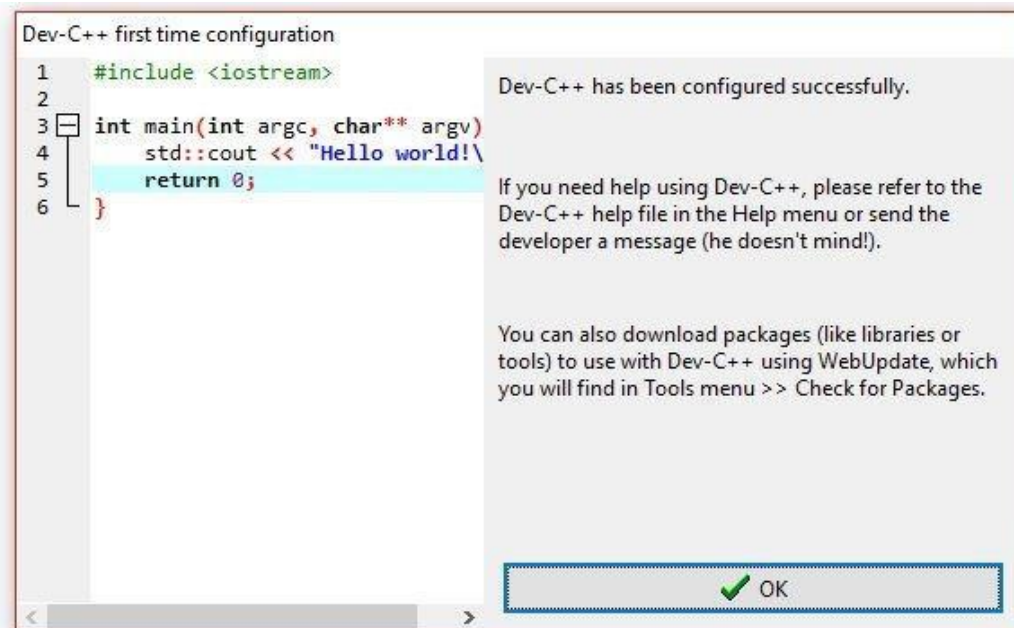


After completing the installation, Dev-C++ will prompt a window with a finish button. Just press the finish button.



After pressing the finish button it asks for language and fonts. If you want to change the language and fonts, then you can change.

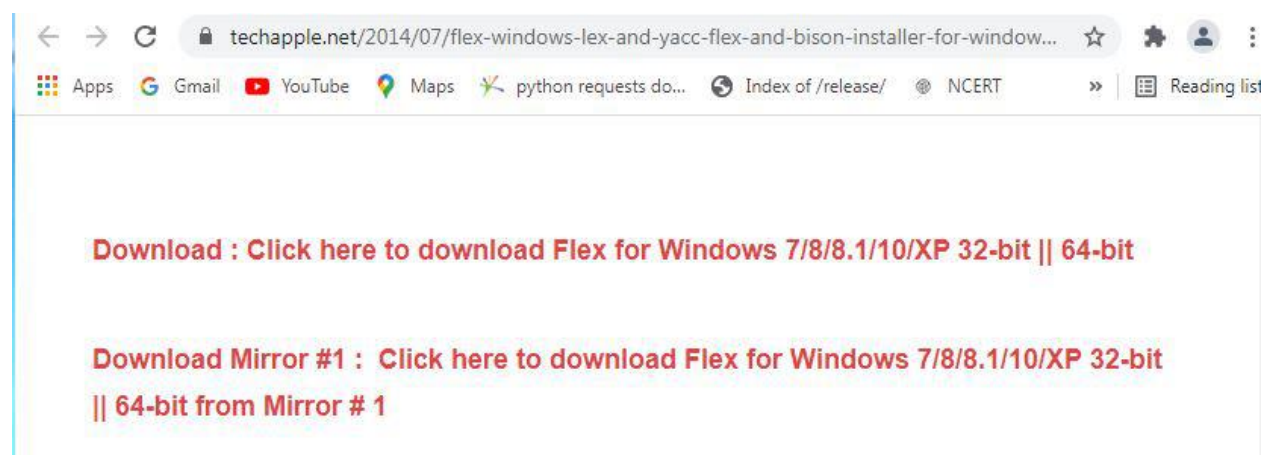




Flex - Lex and Yacc

Step 1 : Install from the above URL

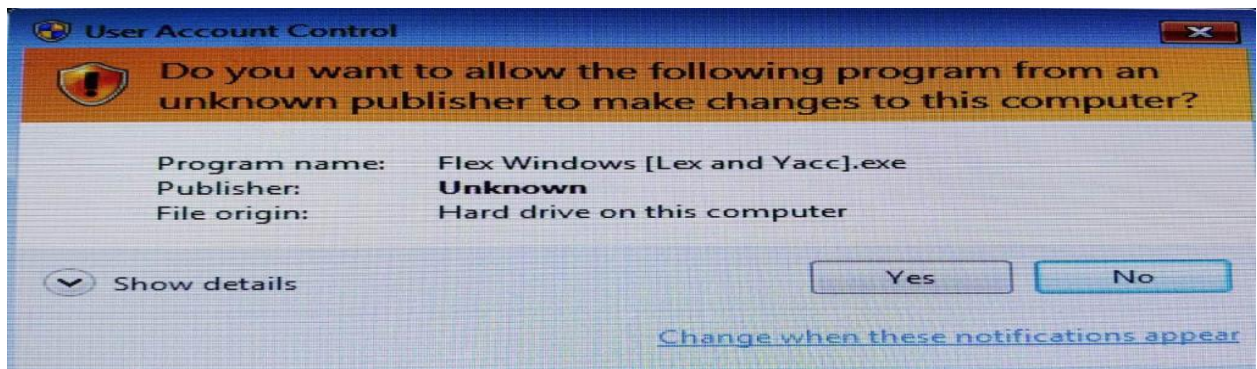
Step 2: On this page, all the features and minimum requirement of the system to install flex is given. Here the download link of the flex program for Windows XP, 7, 8, etc is given. Click on the download link, downloading of the executable file will start shortly. It is a small 30.19 MB file that will hardly take a minute.



Step 3: Now check for the executable file in downloads in your system and run it.



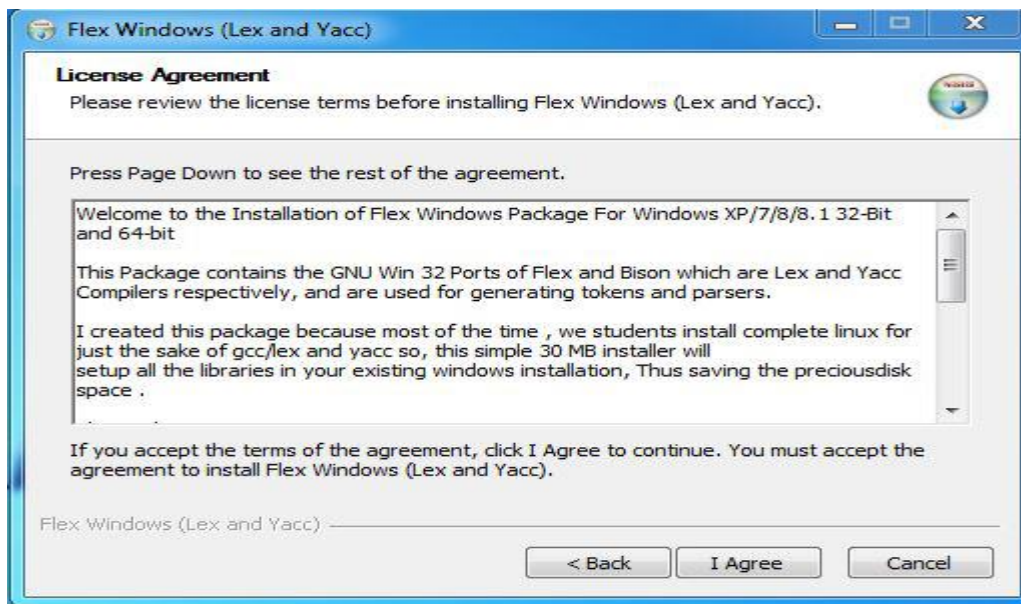
Step 4: It will prompt confirmation to make changes to your system. Click on Yes.



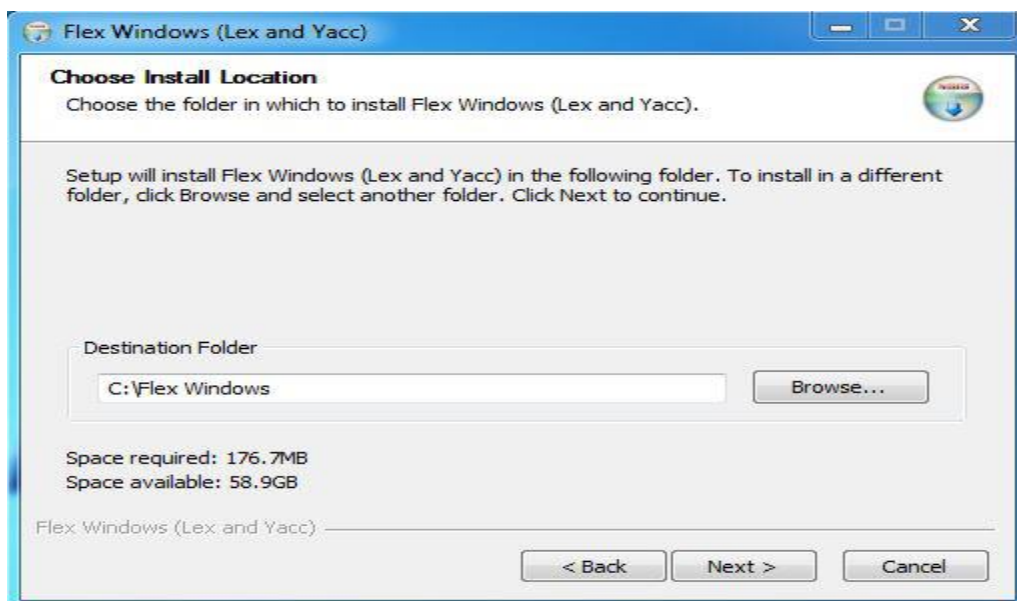
Step 5: Setup screen will appear, click on Next.



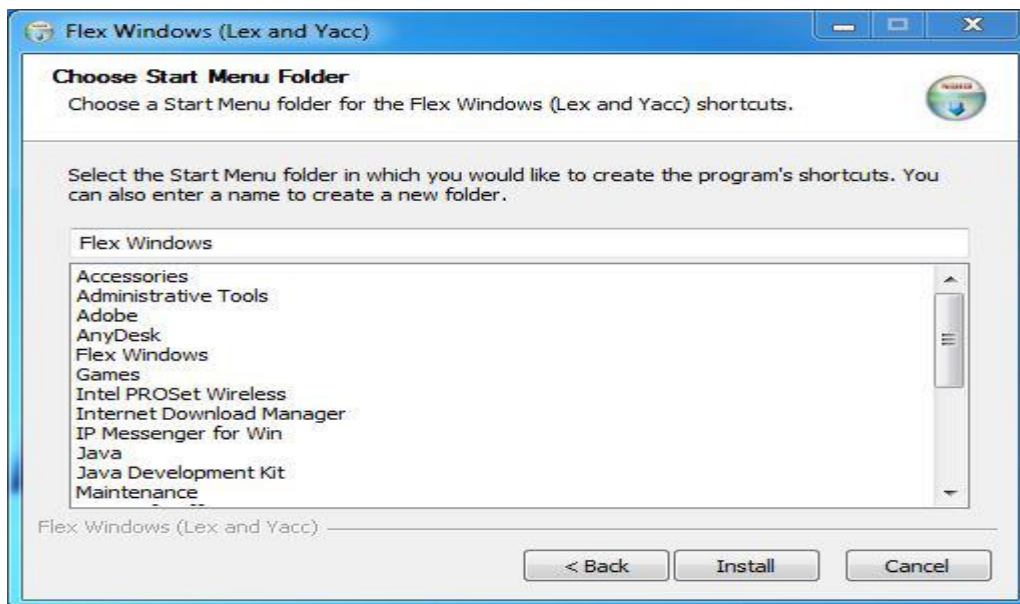
Step 6: The next screen will be of License Agreement, click on I Agree.



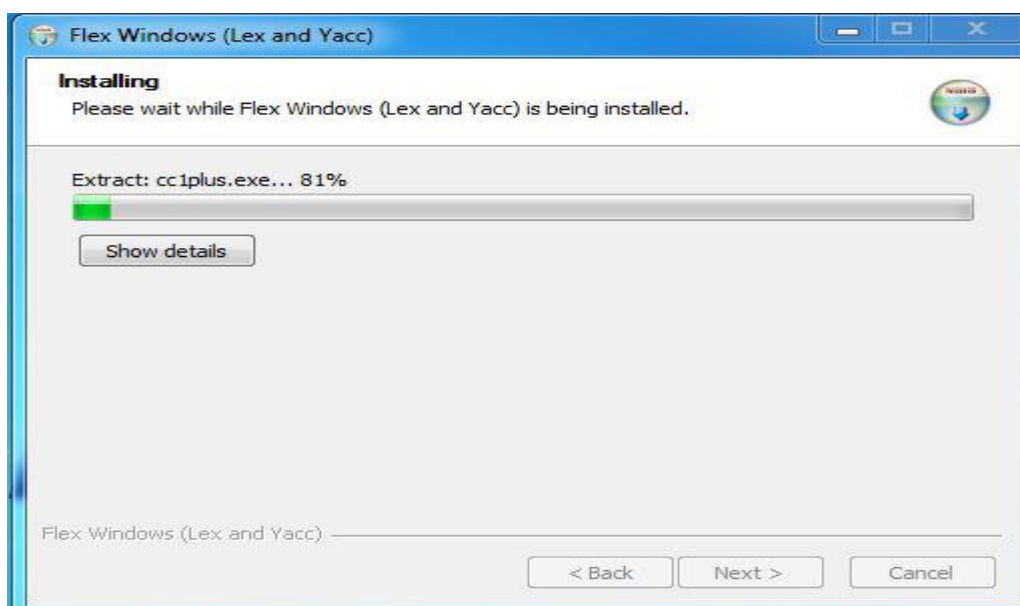
Step 7: The next screen will be of installing location so choose the drive which will have sufficient memory space for installation. It needed only a memory space of 176.7 MB.



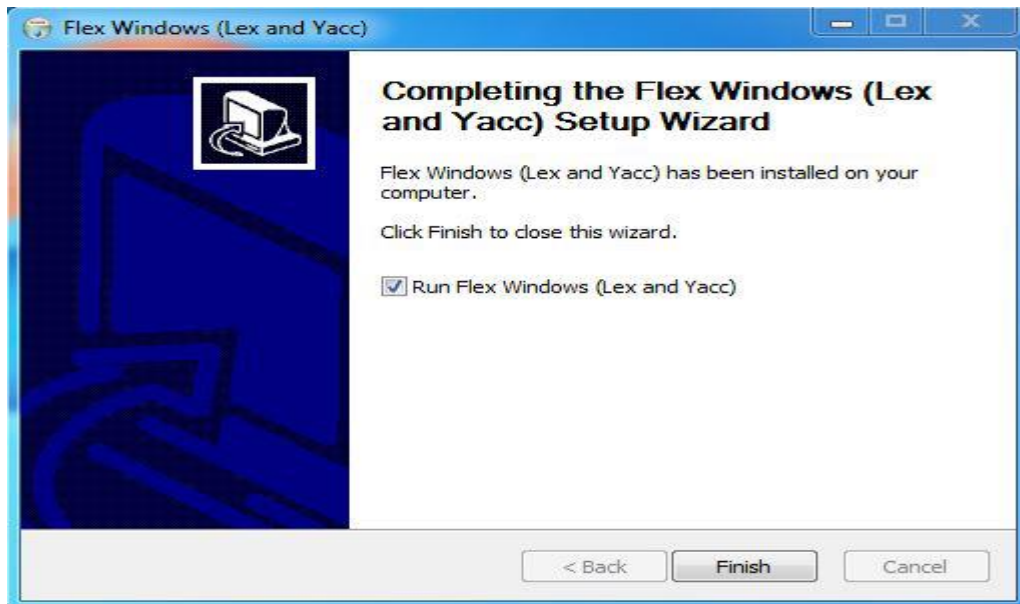
Step 8: Next screen will be of choosing the Start menu folder so don't do anything just click on the Next Button.



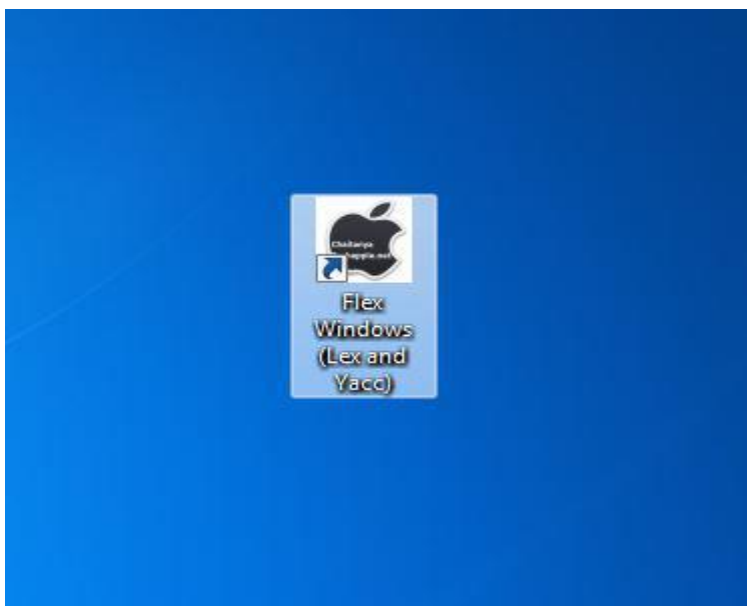
Step 9: After this installation process will start and will hardly take a minute to complete the installation.



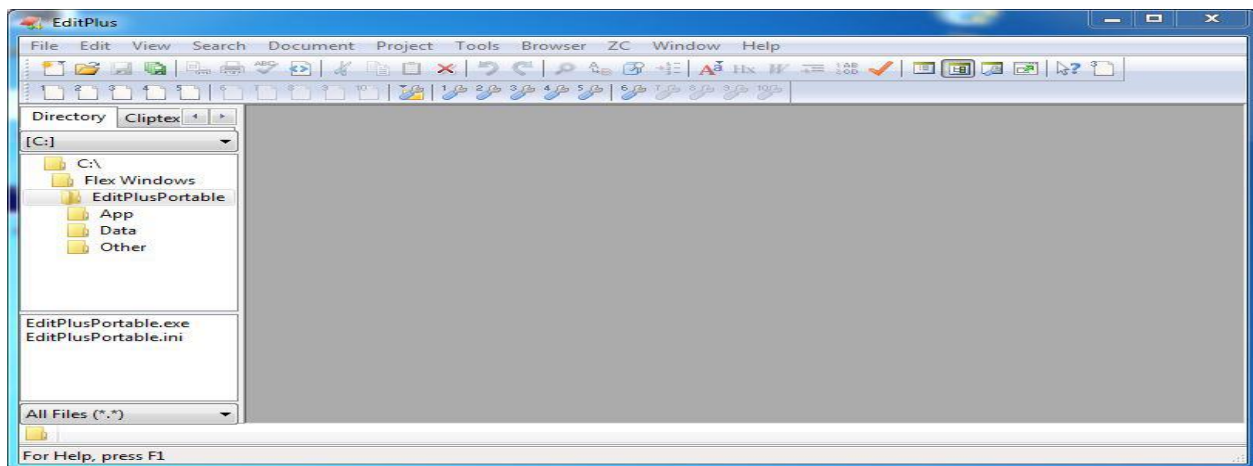
Step 10: Click on Finish after the installation process is complete. Keep the tick mark on the checkbox if you want to run Flex now if not then uncheck it.



Step 11: Flex Windows is successfully installed on the system and an icon is created on the desktop

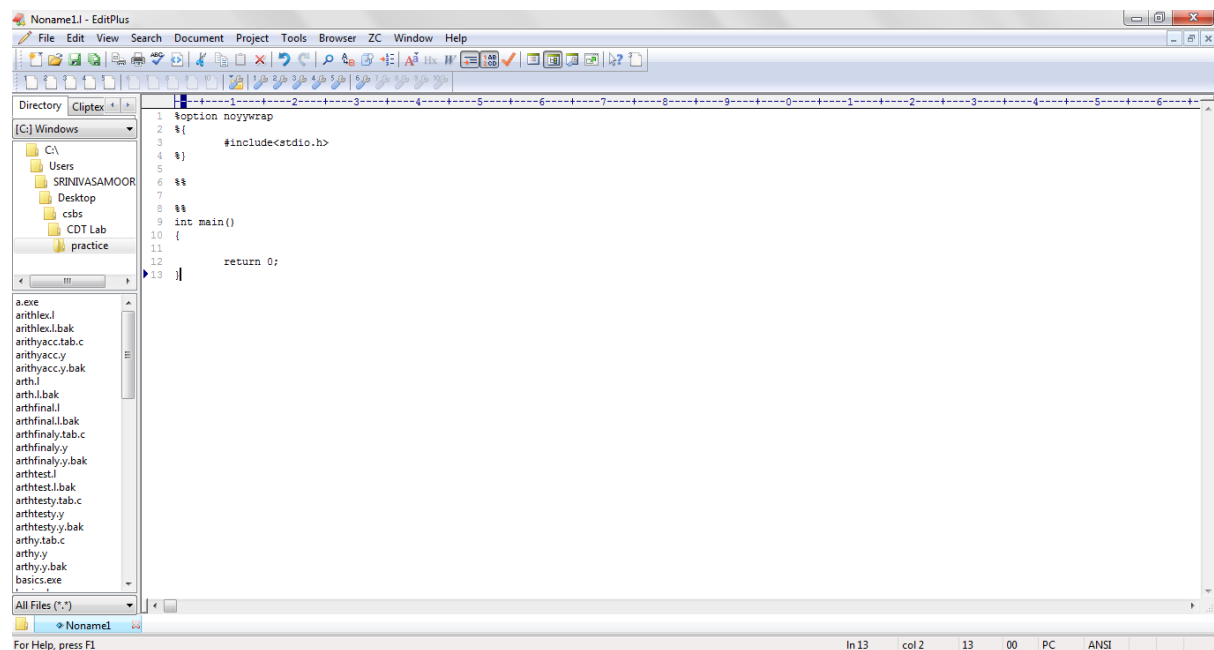


Step 12: Run the software and see the interface.



Flex is now installed on the system

To create a new file file>new>lex then the file would be Noname1.l then we can save and use lex file



To create a new yacc file file>new>yacc/bison then the file would be Noname2.y then we can save and use yacc file

