

**LAB SESSION 2: LEXICAL ANALYSIS WITH ERROR REPORTING USING LEX****DATE:****AIM:** To implement a lexical analyzer for a given programming language using Lex tool..

**PROBLEM DEFINITION:** Design and implement a Lexical Analyzer (Lexer) using the Lex tool for a new: **ExtendC programming language**, which extends the standard C language by introducing two additional datatypes:

1. phone: Represents a phone number.
2. email: Represents an email address.

The lexical analyzer should correctly identify and tokenize the following components from an ExtendC source code file:

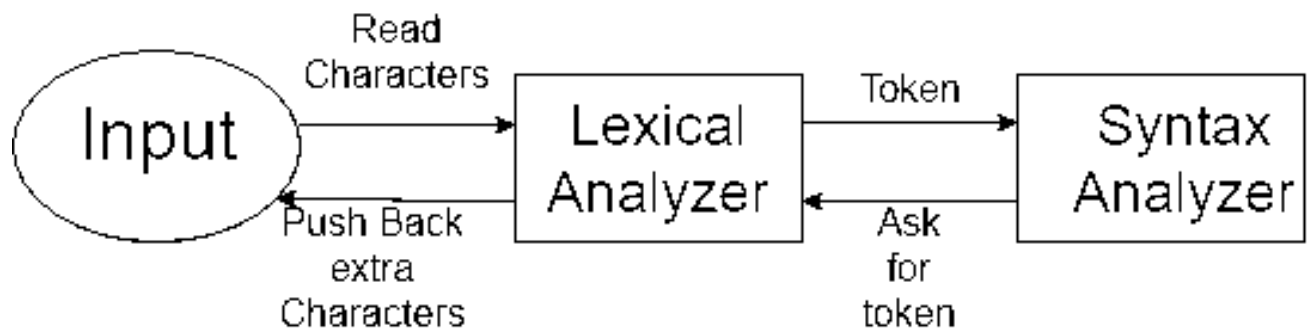
- Keywords: All standard C keywords (e.g., int, char, if, else, for, while, return, etc.), plus the new data types:
  - Phone
  - Email
- Identifiers: Names of variables, functions, etc., following standard C rules.
- Constants:
  - Integer constants
  - Floating-point constants
  - String constants (delimited by double quotes)
  - Character constants (single quotes)
- Special Tokens:
  - Phone Number: Should match the pattern of valid phone numbers (e.g., country code optional, 10 digits). Example valid formats: +91-9876543210, 9876543210
  - Email Address: Should match a valid email pattern as per standard email rules. Example valid formats: [user@example.com](mailto:user@example.com), [first.last@domain.co.in](mailto:first.last@domain.co.in)
  - Operators and Punctuators: +, -, \*, /, =, ==, !=, <, >, <=, >=, {, }, (, ), ,, ,
  - Comments:
    - Single-line comments: // comment text

- Multi-line comments: /\* comment text \*/
  - Whitespace and Newlines: Should be ignored except for counting line numbers for error reporting.

**Expected Output:** For each token identified in the input ExtendC source file, the lexical analyzer should print the token type and its lexeme.

**THEORY:** 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



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, if etc.

Identifier; Examples-Variable name, function name, etc.

Operators; Examples '+', '++', '-' etc.

Separators; Examples ',', ';' etc

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”, “;” .

### **Steps in Lexical Analysis:**

**Input preprocessing:** This stage involves cleaning up the input text and preparing it for lexical analysis. This may include removing comments, whitespace, and other non-essential characters from the input text.

**Tokenization:** This is the process of breaking the input text into a sequence of tokens. This is usually done by matching the characters in the input text against a set of patterns or regular expressions that define the different types of tokens.

**Token classification:** In this stage, the lexer determines the type of each token. For example, in a programming language, the lexer might classify keywords, identifiers, operators, and punctuation symbols as separate token types.

**Token validation:** In this stage, the lexer checks that each token is valid according to the rules of the programming language. For example, it might check that a variable name is a valid identifier, or that an operator has the correct syntax.

**Output generation:** In this final stage, the lexer generates the output of the lexical analysis process, which is typically a list of tokens. This list of tokens can then be passed to the next stage of compilation or interpretation.

The lexical analyzer identifies the error with the help of the automation machine and the grammar of the given language on which it is based like C, C++, and gives row number and column number of the error.

Suppose we pass a statement through lexical analyzer – `a = b + c ;` ; It will generate token sequence like this: `id=id+id;`

**Advantages:**

**Efficiency:** Lexical analysis improves the efficiency of the parsing process because it breaks down the input into smaller, more manageable chunks. This allows the parser to focus on the structure of the code, rather than the individual characters.

**Flexibility:** Lexical analysis allows for the use of keywords and reserved words in programming languages. This makes it easier to create new programming languages and to modify existing ones.

**Error Detection:** The lexical analyzer can detect errors such as misspelled words, missing semicolons, and undefined variables. This can save a lot of time in the debugging process.

**Code Optimization:** Lexical analysis can help optimize code by identifying common patterns and replacing them with more efficient code. This can improve the performance of the program.

**Disadvantages:**

**Complexity:** Lexical analysis can be complex and require a lot of computational power. This can make it difficult to implement in some programming languages.

**Limited Error Detection:** While lexical analysis can detect certain types of errors, it cannot detect all errors. For example, it may not be able to detect logic errors or type errors.

**Increased Code Size:** The addition of keywords and reserved words can increase the size of the code, making it more difficult to read and understand.

**Reduced Flexibility:** The use of keywords and reserved words can also reduce the flexibility of a programming language. It may not be possible to use certain words or phrases in a way that is intuitive to the programmer.

**PROGRAM:****Lex Code:**

```

%{
#include <stdio.h>
#include <string.h>
int lineno = 1;
char comment_buffer[1000]; // Buffer to store multi-line comments
%}

%x COMMENT

%%

("auto"|"break"|"case"|"char"|"const"|"continue"|"default"|"do"|"double"|"phone"|"
email"|"else"|"enum"|"extern"|"float"|"for"|"goto"|"if"|"int"|"long"|"register"|"retur
n"|"short"|"signed"|"sizeof"|"static"|"struct"|"switch"|"typedef"|"union"|"unsigned"
|"void"|"volatile"|"while") { printf("<KEYWORD, %s>\n", yytext); }

\+91-[0-9]{10}      { printf("<PHONE, %s >\n", yytext); }
[0-9]{10}           { printf("<PHONE, %s >\n", yytext); }
[a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,} { printf("<EMAIL, %s >\n", yytext); }

[a-zA-Z_][a-zA-Z0-9_]* { printf("<IDENTIFIER, %s >\n", yytext); }

[0-9]+\.[0-9]+[eE][+-]?[0-9]+ { printf("<FLOAT, %s >\n", yytext); }
[0-9]+[eE][+-]?[0-9]+      { printf("<FLOAT, %s >\n", yytext); }
[0-9]+\.[0-9]+            { printf("<FLOAT, %s >\n", yytext); }
0[xX][0-9a-fA-F]+        { printf("<INTEGER, %s >\n", yytext); }
0[0-7]+                  { printf("<INTEGER, %s >\n", yytext); }
[0-9]+                   { printf("<INTEGER, %s >\n", yytext); }
\"([^\n]|\\.)*\n        { printf("<STRING, %s >\n", yytext); }
\'([^\n]|\\.)\n         { printf("<CHAR, %s >\n", yytext); }

"+="|"-"|"*"|"/"|"%"|"++"|"--"          { printf("<OPERATOR, %s >\n", yytext); }
}
"=="|"!="|"<="|">="|"="|"+"|"-|"*"|" "/"| ">"|"<"      { printf("<OPERATOR, %s >\n",
yytext); }
"{ "|" } "(" ")" ";" ":" "," "[" "]"      { printf("<PUNCTUATOR, %s >\n", yytext); }

"/"/".*"          { printf("<SINGLE_LINE_COMMENT, %s>\n", yytext); }

```

```

^[ \t]*#.*      { printf("<PREPROCESSOR, %s>\n", yytext); }

"/*"          { BEGIN(COMMENT); strcpy(comment_buffer, yytext); }
<COMMENT>"*/"      { strcat(comment_buffer, yytext);
printf("<MULTI_LINE_COMMENT, %s>\n", comment_buffer); BEGIN(INITIAL); }
<COMMENT>\n      { strcat(comment_buffer, yytext); lineno++; }
<COMMENT>.      { strcat(comment_buffer, yytext); }

[ \t\r]+      { /* Ignore whitespace (including CR) */ }
\n            { lineno++; }
[0-9]+[a-zA-Z_][a-zA-Z0-9_]* { printf("ERROR: Invalid identifier starting with digit
'%s' at line %d\n", yytext, lineno); }
.            { printf("ERROR: Unknown symbol '%s' at line %d\n", yytext, lineno); }

%%

int main() {
    yylex();
    return 0;
}

int yywrap() { return 1; }

```

**ex.c file:**

```

// ExtendC Language Comprehensive Test Cases
// Testing all lexical analyzer features

/*
    Multi-line comment
    Testing multi-line comment detection
*/

// 1. Testing Keywords - Standard C Keywords
int main() {
    auto char const continue default do double else enum extern
    float for goto if long register return short signed sizeof
    static struct switch typedef union unsigned void volatile while
    break case
}

```

```
// 2. Testing New ExtendC Keywords
```

```
phone phoneVar;
```

```
email emailVar;
```

```
// 3. Testing Identifiers
```

```
int variable1;
```

```
float _privateVar;
```

```
char userName123;
```

```
int _valid_identifier;
```

```
// 4. Testing Integer Constants
```

```
int num1 = 0;
```

```
int num2 = 123;
```

```
int num3 = 999999;
```

```
// 5. Testing Float Constants
```

```
float pi = 3.14;
```

```
float rate = 0.05;
```

```
float large = 999.999;
```

```
// 6. Testing String Constants
```

```
char* message = "Hello World";
```

```
char* path = "C:\\Users\\Documents\\file.txt";
```

```
char* escaped = "Line1\\nLine2\\tTabbed";
```

```
char* quote = "He said \\\"Hello\\\"";
```

```
// 7. Testing Character Constants
```

```
char ch1 = 'A';
```

```
char ch2 = '5';
```

```
char ch3 = '\\n';
```

```
char ch4 = '\\\"';
```

```
char ch5 = '\\\\';
```

```
// 8. Testing Phone Number Patterns
```

```
phone p1 = 9876543210;    // 10 digits
```

```
phone p2 = +91-9876543210; // With country code
```

```
phone p3 = 1234567890;    // Another valid 10-digit
```

```
// 9. Testing Email Address Patterns
```

```
email e1 = user@example.com;
```

```
email e2 = first.last@domain.co.in;
```

```
email e3 = test123@gmail.com;
email e4 = admin_user@company.org;
email e5 = contact+info@website.net;
```

```
// 10. Testing All Operators
```

```
int a = 5;
int b = 10;
if (a == b) a++;
if (a != b) b--;
if (a < b) a = a + b;
if (a > b) a = a - b;
if (a <= b) a = a * b;
if (a >= b) a = a / b;
```

```
// 11. Testing All Punctuators
```

```
{
    int arr[5];
    func(a, b, c);
    statement1;
    statement2;
}
```

```
// 12. Testing Comments in Different Positions
```

```
int x = 5; // End of line comment
// Comment at beginning of line
/* Inline comment */ int y = 10;
```

```
/*
Multi-line comment
with multiple lines
and various content: symbols @$%^&*()
*/
```

```
// 13. Testing Mixed Complex Expressions
```

```
if ((x + y) >= (a * b)) {
    phone contact = +91-9999888877;
    email notify = admin@system.com;
    char msg = 'Y';
    float result = 123.456;
}
```

```
// 14. Testing Edge Cases
```



```

char empty = "";      // This should cause an error
int negative = -123;   // Testing negative numbers
float scientific = 1.5e10; // This might not be supported
char* emptyStr = "";  // Empty string

```

```

// 15. Testing Error Cases
@ # $ % ^ & * invalid_symbols
invalid-identifier-with-dash
123invalidIdentifier

```

```

// End of test cases

```

### OUTPUT:

```

PS C:\Users\Joseph\Desktop\compiler design\expt2> flex expt2a.l
PS C:\Users\Joseph\Desktop\compiler design\expt2> gcc lex.yy.c -o lex.exe
PS C:\Users\Joseph\Desktop\compiler design\expt2> Get-Content ex.c | ./lex.exe
<SINGLE_LINE_COMMENT, // ExtendC Language Comprehensive Test Cases>
<SINGLE_LINE_COMMENT, // Testing all lexical analyzer features>
<MULTI_LINE_COMMENT, /*
    Multi-line comment
    Testing multi-line comment detection
*/>
<SINGLE_LINE_COMMENT, // 1. Testing Keywords - Standard C Keywords>
<KEYWORD, int>
<IDENTIFIER, main >
<PUNCTUATOR, ( >
<PUNCTUATOR, ) >
<PUNCTUATOR, { >
<KEYWORD, auto>
<KEYWORD, char>
<KEYWORD, const>
<KEYWORD, continue>
<KEYWORD, default>
<KEYWORD, do>
<KEYWORD, double>
<KEYWORD, else>
<KEYWORD, enum>
<KEYWORD, extern>
<KEYWORD, float>
<KEYWORD, for>
<KEYWORD, goto>

```

```
<KEYWORD, if>
<KEYWORD, long>
<KEYWORD, register>
<KEYWORD, return>
<KEYWORD, short>
<KEYWORD, signed>
<KEYWORD, sizeof>
<KEYWORD, static>
<KEYWORD, struct>
<KEYWORD, switch>
<KEYWORD, typedef>
<KEYWORD, union>
<KEYWORD, unsigned>
<KEYWORD, void>
<KEYWORD, volatile>
<KEYWORD, while>
<KEYWORD, break>
<KEYWORD, case>
<PUNCTUATOR, } >
<SINGLE_LINE_COMMENT, // 2. Testing New ExtendC Keywords>
<KEYWORD, phone>
<IDENTIFIER, phoneVar >
<PUNCTUATOR, ; >
<KEYWORD, email>
<IDENTIFIER, emailVar >
<PUNCTUATOR, ; >
<SINGLE_LINE_COMMENT, // 3. Testing Identifiers>
<KEYWORD, int>
<IDENTIFIER, variable1 >
<PUNCTUATOR, ; >
<KEYWORD, float>
<IDENTIFIER, _privateVar >
<PUNCTUATOR, ; >
<KEYWORD, char>
<IDENTIFIER, userName123 >
<PUNCTUATOR, ; >
<KEYWORD, int>
<IDENTIFIER, _valid_identifier >
<PUNCTUATOR, ; >
<SINGLE_LINE_COMMENT, // 4. Testing Integer Constants>
<KEYWORD, int>
<IDENTIFIER, num1 >
```

```
<OPERATOR, = >
<INTEGER, 0 >
<PUNCTUATOR, ; >
<KEYWORD, int>
<IDENTIFIER, num2 >
<OPERATOR, = >
<INTEGER, 123 >
<PUNCTUATOR, ; >
<KEYWORD, int>
<IDENTIFIER, num3 >
<OPERATOR, = >
<INTEGER, 999999 >
<PUNCTUATOR, ; >
<SINGLE_LINE_COMMENT, // 5. Testing Float Constants>
<KEYWORD, float>
<IDENTIFIER, pi >
<OPERATOR, = >
<FLOAT, 3.14 >
<PUNCTUATOR, ; >
<KEYWORD, float>
<IDENTIFIER, rate >
<OPERATOR, = >
<FLOAT, 0.05 >
<PUNCTUATOR, ; >
<KEYWORD, float>
<IDENTIFIER, large >
<OPERATOR, = >
<FLOAT, 999.999 >
<PUNCTUATOR, ; >
<SINGLE_LINE_COMMENT, // 6. Testing String Constants>
<KEYWORD, char>
<OPERATOR, * >
<IDENTIFIER, message >
<OPERATOR, = >
<STRING, "Hello World" >
<PUNCTUATOR, ; >
<KEYWORD, char>
<OPERATOR, * >
<IDENTIFIER, path >
<OPERATOR, = >
<STRING, "C:\\Users\\Documents\\file.txt" >
<PUNCTUATOR, ; >
```

```

<KEYWORD, char>
<OPERATOR, * >
<IDENTIFIER, escaped >
<OPERATOR, = >
<STRING, "Line1\nLine2\tTabbed" >
<PUNCTUATOR, ; >
<KEYWORD, char>
<OPERATOR, * >
<IDENTIFIER, quote >
<OPERATOR, = >
<STRING, "He said \"Hello\"" >
<PUNCTUATOR, ; >
<SINGLE_LINE_COMMENT, // 7. Testing Character Constants >
<KEYWORD, char>
<IDENTIFIER, ch1 >
<OPERATOR, = >
<CHAR, 'A' >
<PUNCTUATOR, ; >
<KEYWORD, char>
<IDENTIFIER, ch2 >
<OPERATOR, = >
<CHAR, '5' >
<PUNCTUATOR, ; >
<KEYWORD, char>
<IDENTIFIER, ch3 >
<OPERATOR, = >
<CHAR, '\n' >
<PUNCTUATOR, ; >
<KEYWORD, char>
<IDENTIFIER, ch4 >
<OPERATOR, = >
<CHAR, '\"' >
<PUNCTUATOR, ; >
<KEYWORD, char>
<IDENTIFIER, ch5 >
<OPERATOR, = >
<CHAR, '\\' >
<PUNCTUATOR, ; >
<SINGLE_LINE_COMMENT, // 8. Testing Phone Number Patterns>
<KEYWORD, phone>
<IDENTIFIER, p1 >
<OPERATOR, = >

```

```
<PHONE, 9876543210 >
<PUNCTUATOR, ; >
<SINGLE_LINE_COMMENT, // 10 digits>
<KEYWORD, phone>
<IDENTIFIER, p2 >
<OPERATOR, = >
<PHONE, +91-9876543210 >
<PUNCTUATOR, ; >
<SINGLE_LINE_COMMENT, // With country code>
<KEYWORD, phone>
<IDENTIFIER, p3 >
<OPERATOR, = >
<PHONE, 1234567890 >
<PUNCTUATOR, ; >
<SINGLE_LINE_COMMENT, // Another valid 10-digit>
<SINGLE_LINE_COMMENT, // 9. Testing Email Address Patterns>
<KEYWORD, email>
<IDENTIFIER, e1 >
<OPERATOR, = >
<EMAIL, user@example.com >
<PUNCTUATOR, ; >
<KEYWORD, email>
<IDENTIFIER, e2 >
<OPERATOR, = >
<EMAIL, first.last@domain.co.in >
<PUNCTUATOR, ; >
<KEYWORD, email>
<IDENTIFIER, e3 >
<OPERATOR, = >
<EMAIL, test123@gmail.com >
<PUNCTUATOR, ; >
<KEYWORD, email>
<IDENTIFIER, e4 >
<OPERATOR, = >
<EMAIL, admin_user@company.org >
<PUNCTUATOR, ; >
<KEYWORD, email>
<IDENTIFIER, e5 >
<OPERATOR, = >
<EMAIL, contact+info@website.net >
<PUNCTUATOR, ; >
<SINGLE_LINE_COMMENT, // 10. Testing All Operators>
```

<KEYWORD, int>  
<IDENTIFIER, a >  
<OPERATOR, = >  
<INTEGER, 5 >  
<PUNCTUATOR, ; >  
<KEYWORD, int>  
<IDENTIFIER, b >  
<OPERATOR, = >  
<INTEGER, 10 >  
<PUNCTUATOR, ; >  
<KEYWORD, if>  
<PUNCTUATOR, ( >  
<IDENTIFIER, a >  
<OPERATOR, == >  
<IDENTIFIER, b >  
<PUNCTUATOR, ) >  
<IDENTIFIER, a >  
<OPERATOR, ++ >  
<PUNCTUATOR, ; >  
<KEYWORD, if>  
<PUNCTUATOR, ( >  
<IDENTIFIER, a >  
<OPERATOR, != >  
<IDENTIFIER, b >  
<PUNCTUATOR, ) >  
<IDENTIFIER, b >  
<OPERATOR, -- >  
<PUNCTUATOR, ; >  
<KEYWORD, if>  
<PUNCTUATOR, ( >  
<IDENTIFIER, a >  
<OPERATOR, < >  
<IDENTIFIER, b >  
<PUNCTUATOR, ) >  
<IDENTIFIER, a >  
<OPERATOR, = >  
<IDENTIFIER, a >  
<OPERATOR, + >  
<IDENTIFIER, b >  
<PUNCTUATOR, ; >  
<KEYWORD, if>  
<PUNCTUATOR, ( >

<IDENTIFIER, a >  
<OPERATOR, > >  
<IDENTIFIER, b >  
<PUNCTUATOR, ) >  
<IDENTIFIER, a >  
<OPERATOR, = >  
<IDENTIFIER, a >  
<OPERATOR, - >  
<IDENTIFIER, b >  
<PUNCTUATOR, ; >  
<KEYWORD, if>  
<PUNCTUATOR, ( >  
<IDENTIFIER, a >  
<OPERATOR, <= >  
<IDENTIFIER, b >  
<PUNCTUATOR, ) >  
<IDENTIFIER, a >  
<OPERATOR, = >  
<IDENTIFIER, a >  
<OPERATOR, \* >  
<IDENTIFIER, b >  
<PUNCTUATOR, ; >  
<KEYWORD, if>  
<PUNCTUATOR, ( >  
<IDENTIFIER, a >  
<OPERATOR, >= >  
<IDENTIFIER, b >  
<PUNCTUATOR, ) >  
<IDENTIFIER, a >  
<OPERATOR, = >  
<IDENTIFIER, a >  
<OPERATOR, / >  
<IDENTIFIER, b >  
<PUNCTUATOR, ; >  
<SINGLE\_LINE\_COMMENT, // 11. Testing All Punctuators>  
<PUNCTUATOR, { >  
<KEYWORD, int>  
<IDENTIFIER, arr >  
<PUNCTUATOR, [ >  
<INTEGER, 5 >  
<PUNCTUATOR, ] >  
<PUNCTUATOR, ; >

```

<IDENTIFIER, func >
<PUNCTUATOR, ( >
<IDENTIFIER, a >
<PUNCTUATOR, , >
<IDENTIFIER, b >
<PUNCTUATOR, , >
<IDENTIFIER, c >
<PUNCTUATOR, ) >
<PUNCTUATOR, ; >
<IDENTIFIER, statement1 >
<PUNCTUATOR, ; >
<IDENTIFIER, statement2 >
<PUNCTUATOR, ; >
<PUNCTUATOR, } >
<SINGLE_LINE_COMMENT, // 12. Testing Comments in Different Positions>
<KEYWORD, int>
<IDENTIFIER, x >
<OPERATOR, = >
<INTEGER, 5 >
<PUNCTUATOR, ; >
<SINGLE_LINE_COMMENT, // End of line comment>
<SINGLE_LINE_COMMENT, // Comment at beginning of line>
<MULTI_LINE_COMMENT, /* Inline comment */>
<KEYWORD, int>
<IDENTIFIER, y >
<OPERATOR, = >
<INTEGER, 10 >
<PUNCTUATOR, ; >
<MULTI_LINE_COMMENT, /*
Multi-line comment
with multiple lines
and various content: symbols @$%^&*()
*/>
<SINGLE_LINE_COMMENT, // 13. Testing Mixed Complex Expressions>
<KEYWORD, if>
<PUNCTUATOR, ( >
<PUNCTUATOR, ( >
<IDENTIFIER, x >
<OPERATOR, + >
<IDENTIFIER, y >
<PUNCTUATOR, ) >
<OPERATOR, >= >

```



```

<PUNCTUATOR, ( >
<IDENTIFIER, a >
<OPERATOR, * >
<IDENTIFIER, b >
<PUNCTUATOR, ) >
<PUNCTUATOR, ) >
<PUNCTUATOR, { >
<KEYWORD, phone>
<IDENTIFIER, contact >
<OPERATOR, = >
<PHONE, +91-9999888877 >
<PUNCTUATOR, ; >
<KEYWORD, email>
<IDENTIFIER, notify >
<OPERATOR, = >
<EMAIL, admin@system.com >
<PUNCTUATOR, ; >
<KEYWORD, char>
<IDENTIFIER, msg >
<OPERATOR, = >
<CHAR, 'Y' >
<PUNCTUATOR, ; >
<KEYWORD, float>
<IDENTIFIER, result >
<OPERATOR, = >
<FLOAT, 123.456 >
<PUNCTUATOR, ; >
<PUNCTUATOR, } >
<SINGLE_LINE_COMMENT, // 14. Testing Edge Cases>
<KEYWORD, char>
<IDENTIFIER, empty >
<OPERATOR, = >
ERROR: Unknown symbol "" at line 100
ERROR: Unknown symbol "" at line 100
<PUNCTUATOR, ; >
<SINGLE_LINE_COMMENT, // This should cause an error>
<KEYWORD, int>
<IDENTIFIER, negative >
<OPERATOR, = >
<OPERATOR, - >
<INTEGER, 123 >
<PUNCTUATOR, ; >

```

```

<SINGLE_LINE_COMMENT, // Testing negative numbers>
<KEYWORD, float>
<IDENTIFIER, scientific >
<OPERATOR, = >
<FLOAT, 1.5e10 >
<PUNCTUATOR, ; >
<SINGLE_LINE_COMMENT, // This might not be supported>
<KEYWORD, char>
<OPERATOR, * >
<IDENTIFIER, emptyStr >
<OPERATOR, = >
<STRING, "" >
<PUNCTUATOR, ; >
<SINGLE_LINE_COMMENT, // Empty string>
<SINGLE_LINE_COMMENT, // 15. Testing Error Cases>
ERROR: Unknown symbol '@' at line 106
ERROR: Unknown symbol '#' at line 106
ERROR: Unknown symbol '$' at line 106
ERROR: Unknown symbol '%' at line 106
ERROR: Unknown symbol '^' at line 106
ERROR: Unknown symbol '&' at line 106
<OPERATOR, * >
<IDENTIFIER, invalid_symbols >
<IDENTIFIER, invalid >
<OPERATOR, - >
<IDENTIFIER, identifier >
<OPERATOR, - >
<IDENTIFIER, with >
<OPERATOR, - >
<IDENTIFIER, dash >
ERROR: Invalid identifier starting with digit '123invalidIdentifier' at line 108
<SINGLE_LINE_COMMENT, // End of test cases>
PS C:\Users\Joseph\Desktop\compiler design\expt2>

```

**CONCLUSION:**

In this experiment, we implemented a Lexical Analyzer for the extended C language (ExtendC) using Lex. It successfully identified keywords, identifiers, constants, operators, punctuators, and the new data types phone and email, recognizing valid phone numbers and email addresses as special tokens while reporting errors for invalid symbols with line numbers.

We learned how Lex converts regular expressions into finite state automata (FSA) for token recognition, how to extend a language with new tokens and patterns, and how to classify tokens effectively. The experiment also emphasized handling whitespace, line numbers, and errors, highlighting the importance of lexical analysis as the foundation for syntax and semantic analysis. Additionally, practical skills like processing source code files using file redirection were reinforced.