North Western University

Compiler Design
(User manual)

Course Code: CSE-4104

Developed By:

Afia Farzana Maliha

20201109010

Sumaiya Juthi

20201115010

Department: CSE

4th year, 1st semester

Contents

Abstract:

The purpose of this lab project was to design and implement a lexical analyzer, also known as a lexer, for a programming language. The lexer is an essential component of a compiler or interpreter, responsible for breaking down the source code into tokens that can be processed further. This report discusses the design choices, implementation details, and showing symbol table of this project.

Introduction:

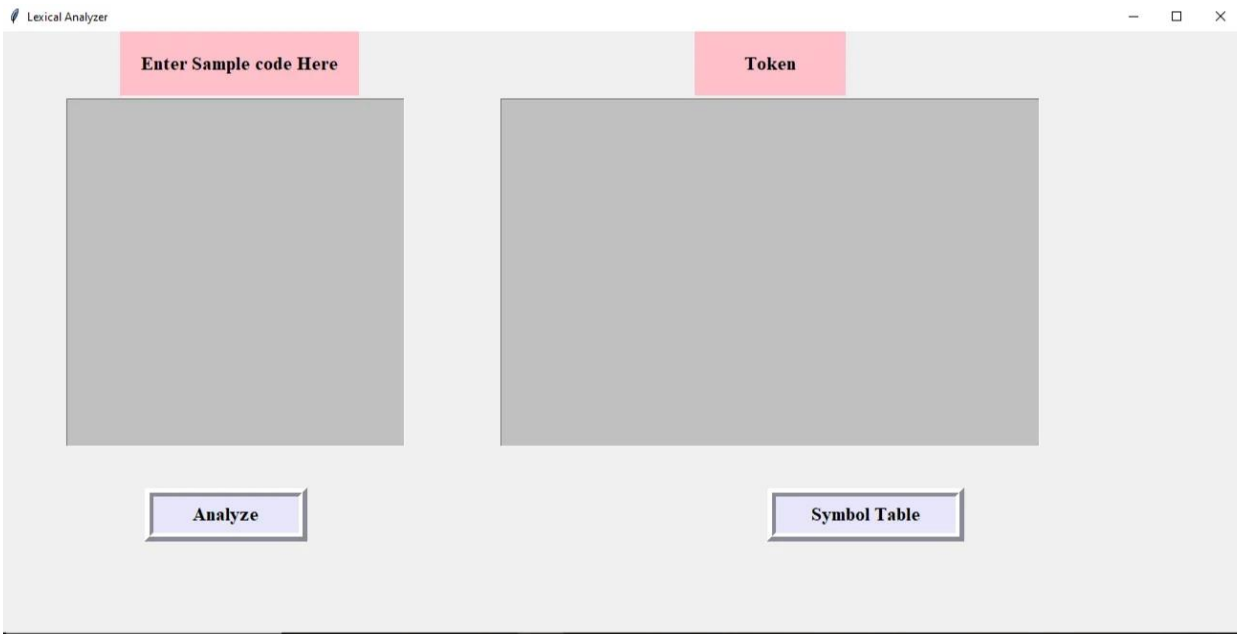
The lexical analyzer is the first phase of a compiler It takes modified source code from language preprocessors that are written in the form of sentences. The lexical analyzer breaks these syntaxes into a series of tokens, by removing any white space or comments in the source code. These tokens include keywords, identifiers, and punctuation symbols.

Objectives:

The objective of this project is to create a program that can analyze the source code of a programming language and break it down into individual tokens. The program should accurately recognize and categorize tokens such as keywords, identifiers, literals, operators, and punctuation symbols. It should also handle errors and provide informative error messages. The project aims to create an efficient and reliable tool that can integrate with other components of a compiler or interpreter for further analysis and processing of the source code.

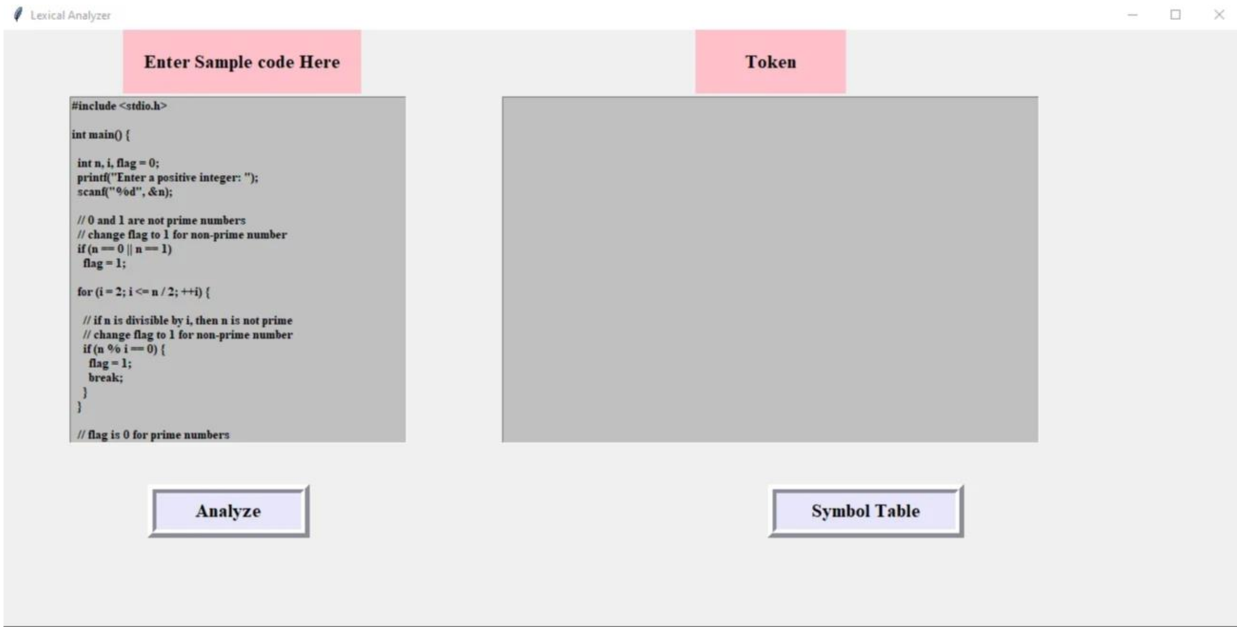
Design And Implementation:

The lexer was implemented using the Python programming language and special use of python tkinter library. The design involved creating a set of c language code to match the different token types and then applying them to the source code in a sequential manner. The lexer followed the longest match rule, where it selects the longest matching token from the input stream. There are thress steps are input source code, show tokenization and show symbol table.



Project Interface

Input Buffering:

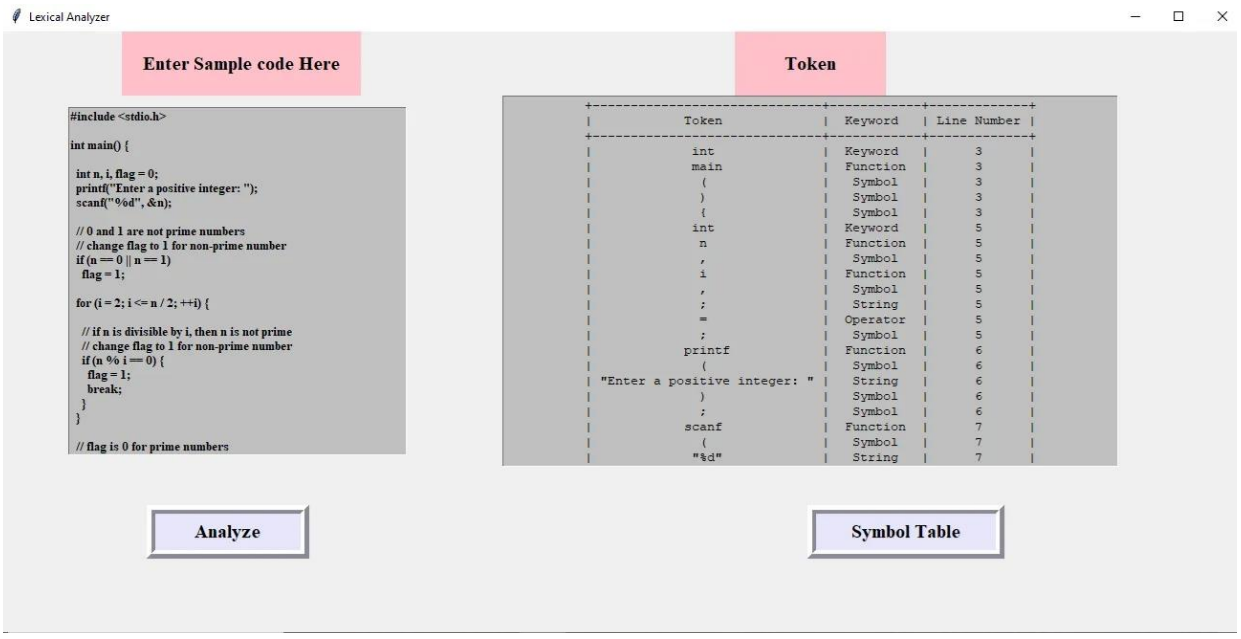


```c
#include <stdio.h>

int main() {

 int n, i, flag = 0;
 printf("Enter a positive integer: ");
 scanf("%d", &n);

 // 0 and 1 are not prime numbers
 // change flag to 1 for non-prime number
 if (n == 0 || n == 1)
  flag = 1;

 for (i = 2; i <= n / 2; ++i) {

  // if n is divisible by i, then n is not prime
  // change flag to 1 for non-prime number
  if (n % i == 0) {
   flag = 1;
   break;
  }
 }

 // flag is 0 for prime numbers
```

The source code has been input in the input field. Only C language is supported as source code.

Clear source code button is used for clear the input field.

Tokenization:

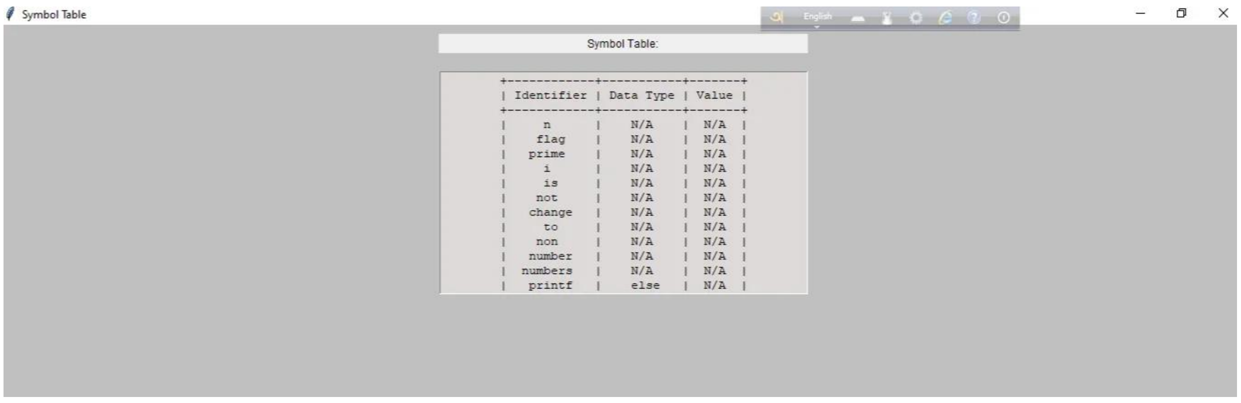


```
#include <stdio.h>

int main() {

 int n, i, flag = 0;
 printf("Enter a positive integer: ");
 scanf("%d", &n);

 // 0 and 1 are not prime numbers
 // change flag to 1 for non-prime number
 if (n == 0 || n == 1)
  flag = 1;

 for (i = 2; i <= n / 2; ++i) {

  // if n is divisible by i, then n is not prime
  // change flag to 1 for non-prime number
  if (n % i == 0) {
   flag = 1;
   break;
  }
 }

 // flag is 0 for prime numbers
```

```
+------------------------------+------------+-------------+
|            Token             |  Keyword   | Line Number |
+------------------------------+------------+-------------+
|             int              |  Keyword   |      3      |
|             main             |  Function  |      3      |
|              (               |   Symbol   |      3      |
|              )               |   Symbol   |      3      |
|              {               |   Symbol   |      3      |
|             int              |  Keyword   |      5      |
|              n               |  Function  |      5      |
|              ,               |   Symbol   |      5      |
|              i               |  Function  |      5      |
|              ,               |   Symbol   |      5      |
|              ;               |   String   |      5      |
|              =               |  Operator  |      5      |
|              ;               |   Symbol   |      5      |
|            printf            |  Function  |      6      |
|              (               |   Symbol   |      6      |
| "Enter a positive integer: " |   String   |      6      |
|              )               |   Symbol   |      6      |
|              ;               |   Symbol   |      6      |
|            scanf             |  Function  |      7      |
|              (               |   Symbol   |      7      |
|             "%d"             |   String   |      7      |
```

The lexer iterated over the input buffer character by character, recognizing and categorizing each token it encountered. Source code were utilized to define the patterns for various tokens, such as keywords, identifiers, number, strings symbols etc. Whole process handled after clicking analyze button. And there is clear tokenize button for clear the tokenize table. The lexer maintained a symbol table to store and manage identifiers encountered during tokenization.

Symbol Table:



```
+------------+-----------+-------+
| Identifier | Data Type | Value |
+------------+-----------+-------+
|     n      |    N/A    |  N/A  |
|    flag    |    N/A    |  N/A  |
|   prime    |    N/A    |  N/A  |
|     i      |    N/A    |  N/A  |
|     is     |    N/A    |  N/A  |
|    not     |    N/A    |  N/A  |
|   change   |    N/A    |  N/A  |
|     to     |    N/A    |  N/A  |
|    non     |    N/A    |  N/A  |
|   number   |    N/A    |  N/A  |
|  numbers   |    N/A    |  N/A  |
|   printf   |    else   |  N/A  |
```

Symbol is used for hold the identifiers of the source code. So in this symbol table hold the identifiers and the value of each identifiers.

Conclusion:

In this project successfully implemented a lexer for a programming language. The design and implementation of the lexer allowed for accurate and efficient token recognition. The project achieved its objective of breaking down the source code into meaningful tokens, laying the foundation for subsequent phases of the compiler.

Thanks to

Md. Shymon Islam

Lecturer

Department Of CSE

North Western University Khulna