# Universidad de Guadalajara

# Centro Universitario de Ciencias Exactas e Ingenierías DIVISIÓN DE TECNOLOGÍAS PARA LA INTEGRACIÓN CIBER-HUMANA

Departamento de Ciencias Computacionales



# Seminario de Solución de Problemas de Traductores de Lenguajes II Sección: D06

Subproducto integrador 1. Analizador Léxico

#### Presenta:

215638591 - Avila Guerrero Ismael Guadalupe 215769734 - Hernandez Chavez Isaac Alberto 219747646 - Martinez Zepeda Diego Aalberto

Profesor: Roberto Patiño Ruiz

Fecha de entrega: 23 de septiembre de 2023

# ÍNDICE

OBJETIVO	2
DESARROLLO	3
Requerimientos Funcionales	3
Requerimientos no Funcionales	5
Gramática	7
Autómata	8
Interfaz Gráfica	9
Código Fuente	10
Pruebas de Ejecución	18
CONCLUSIONES PERSONALES	19
REFERENCIAS BIBLIOGRÁFICAS	20

## **OBJETIVO**

Aplicar los fundamentos básicos de los compiladores, para desarrollar un analizador léxico capaz de leer un flujo de caracteres de entrada (símbolos léxicos) y transformarlo en una secuencia de componentes léxicos que posteriormente utilizará el analizador sintáctico.

## **DESARROLLO**

Para el desarrollo del analizador léxico se tomaron en cuenta los siguientes requerimientos funcionales y no funcionales:

# **Requerimientos Funcionales**

Número de requerimiento	1		
Nombre del requerimiento	Permitir la entrada de caracteres para dar el proceso de análisis		
Tipo	☑ <del>Requerimiento</del>		☐ Restricción
Prioridad del requerimiento	☑ Alta/Esencial	☐ Media/Dese ado	☐ Baja/Opcion al
Número de requerimiento	2		
Nombre del requerimiento	Eliminar los blancos o comentarios		
Tipo	☑ <del>Requerimiento</del>		☐ Restricción
Prioridad del requerimiento	☑ <del>Alta/Esencial</del>	☐ Media/Dese ado	□ Baja/Opcion al

Número de requerimiento	3		
Nombre del requerimiento	Hacer el proceso de verificación del análisis léxico con la cadena ingresada por el usuario		
Tipo	☑ <del>Requerimiento</del>		☐ Restricción
Prioridad del requerimiento	☑ <del>Alta/Esencial</del>	☐ Media/Dese ado	□ Baja/Opcion al
Número de requerimiento	4		
Nombre del requerimiento	Mostrar en pantalla los token encontrados en la cadena ingresada por el usuario		
Tipo	☑ <del>Requerimiento</del>		☐ Restricción
Prioridad del requerimiento	☑ <del>Alta/Esencial</del>	☐ Media/Dese ado	□ Baja/Opcion al

# Requerimientos no Funcionales

Número de requerimiento	1		
Nombre del requerimiento	Mostrar en pantalla la ejecución del programa exitósamente		
Tipo	☑ <del>Requerimiento</del>		☐ Restricción
Prioridad del requerimiento	☑ <del>Alta/Esencial</del>	☐ Media/Dese ado	☐ Baja/Opcion al
Número de requerimiento	2		
Nombre del requerimiento	Usar una interfaz gráfica que permita al usuario hacer uso de la herramienta sin conocerla previamente		
Tipo	☑ <del>Requerimiento</del>		☐ Restricción
Prioridad del requerimiento	☑ Alta/Esencial	☐ Media/Dese ado	□ Baja/Opcion al
Número de requerimiento		3	
Nombre del requerimiento	Usar una interfaz gráfica que permita al usuario sentir que hay una linealidad en los pasos a seguir para usar la herramienta		

Tipo	☑ <del>Requerimiento</del>		☐ Restricción
Prioridad del requerimiento	☐ Alta/Esencial	☑ <del>Media/Dese</del> <del>ado</del>	☐ Baja/Opcion al

## Gramática

La gramática utilizada para reconocer los símbolos es la siguiente:

Token	Tipo	Lexema
Identificador	0	
Número	1	
Punto y coma	2	;
Coma	3	,
Paréntisis Izq	4	(
Paréntesis Der	5	)
Llave Izq	6	{
Llave Der	7	}

Token	Tipo	Lexema
Signo de Pesos	8	\$
Operador Relacional	9	!=, ==, <, <=, >, >=
Asignación	10	=
Operador Lógico	11	, &&
Operador Suma	12	+, -
Operador Multiplicación	13	*, /
Tipo de Dato	14	int, float
Palabra Reservada	15	if, else, while, return

Imagen 1. Gramática.

## **Autómata**

El autómata que utilizamos para el desarrollo del analizador léxico es el siguiente:

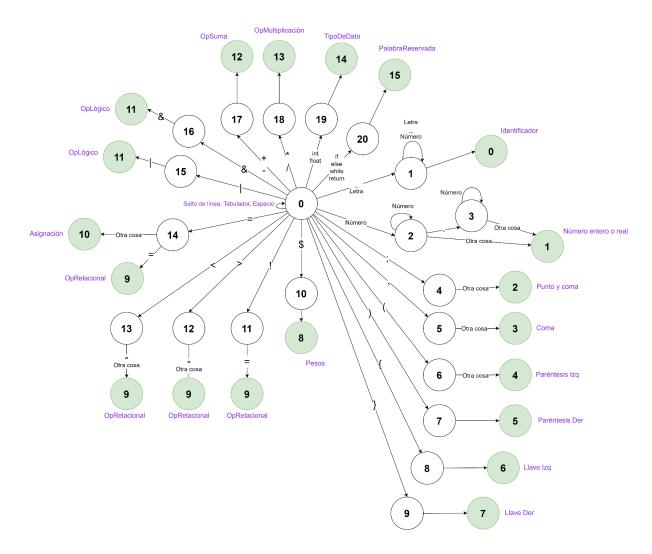


Imagen 2. Autómata.

### Interfaz Gráfica

El analizador léxico utilizamos el lenguaje de programación C#, debido a que es un lenguaje de alto nivel, es un lenguaje orientado a objetos y cuenta con una sintaxis muy familiar para nosotros, ya que es muy similar a lenguajes como C++ y Java.

Otra de las razones para utilizar este lenguaje es que el framework de .NET tiene soporte para los Windows Forms, que son elementos de interfaz gráfica para crear aplicaciones de escritorio nativas para Windows.

La interfaz gráfica desarrollada está pensada para ser muy intuitiva para el usuario. Cuenta con un input en el que se puede ingresar texto, el cual debe ser código fuente. El botón con la leyenda "Analizador Léxico" toma el texto ingresado, lo analiza y crea una lista de componentes léxicos que se muestra en el listado de la parte inferior.

El diseño implementado es el siguiente:



Imagen 3. Interfaz gráfica de usuario.

## Código Fuente

El analizador léxico se dividió en varias clases para facilitar el desarrollo. La primera de ellas es la clase LexicalComponents, que guarda las propiedades de los componentes léxicos reconocidos por el analizador.

La clase cuenta con un constructor que recibe como parámetro los valores para inicializar el componente léxico. Además, tiene implementadas las propiedades para acceder y modificar los atributos de cada objeto de esta clase.

El código de esta clase es el siguiente:

```
internal class LexicalComponent
 2
          string lexeme;
          string token;
 5
          int number;
 6
          public LexicalComponent() { }
 8
 9
          public LexicalComponent(string lexeme, string token, int number)
11
              this.lexeme = lexeme;
              this.token = token;
13
              this.number = number;
14
15
          public string Lexeme
16
17
              get { return lexeme; }
19
              set { lexeme = value; }
20
21
          public string Token
22
23
24
              get { return token; }
25
              set { token = value; }
27
          public int Numbe
28
              get { return number; }
30
31
              set { number = value; }
          }
32
      }
33
```

Imagen 4. Código de la clase LexicalComponent.

El analizador léxico se implementó en una clase llamada Analyzers. La gramática se cargó en un enum para facilitar su uso a lo largo del análisis.

```
1 enum Grammar
      Identifier = 0,
      Number = 1,
     Semicolon = 2,
 5
 6 Comma = 3,
     LeftParenthesis = 4,
8 RightParenthesis = 5,
9 LeftCurlyBrace = 6,
10 RightCurlyBrace = 7,
DollarSing = 8,
RelationalOperator = 9,
   AssignmentOperator = 10,
LogicOperator = 11,
15
     AdditionOperator = 12,
16
     MultiplicationOperator = 13,
17
     DataType = 14,
     KeyWord = 15
18
19 }
```

Imagen 5. Gramática cargada en un enum.

El método LexicalAnalyze es un método estático que recibe como parámetros un string y la lista de elementos léxicos. El string es una línea de entrada proporcionada por el usuario para el analizador.

El analizador se implementó con una robusta estructura de switch-case. El analizador recorre caracter por caracter el string recibido. Se inicializan una serie de variables que son usadas para guardar el estado actual, el lexema que se va construyendo y el token correspondiente al lexema.

Conforme se va reconociendo un lexema, se agrega a la lista de elementos léxicos y se reinician las variables antes mencionadas. El código fuente del analizador léxico se detalla a continuación.

El case 0 es el más extenso de todos, ya que es donde inicia el reconocimiento de los lexemas (ver Imagen 2).

```
. . .
        1 case 0:
2  if (Char.IsLetter(line[i]) || line[i] = '_') {
                        state = 1;
lexeme += line[i];
token = Grammar.Identifier.ToString();
                       if (!Char.IsDigit(!ine[i + 1]) 66 !Char.IsLetter(!ine[i + 1]) 66 !ine[i + 1] ≠ '_') {
   if (!exeme = "ini" || lexeme = "float")
   elements.Add(new texticalComponent(!exeme, Grammar.DataType.ToString(), (int)Grammar.DataType));
                                else if (lexeme = "if" || lexeme = "else" || lexeme = "while" || lexeme = "return")
    elements.Add(new LexicalComponent(lexeme, Grammar.KeyWord.ToString(), (int)Grammar.KeyWord));
                                else
    elements.Add(new LexicalComponent(lexeme, token, (int)Grammar.Identifier));
                               state = 0; lexeme = ""; token = "";
              break;
}
else if (Char.IsDigit(line[i])) {
    state = 2;
    lexem = + inne[i];
    token = Grammar.Number.ToString();
    break;
              break;
} else f(ine[i] = ';') {
    lexeme += line[i];
    elseents.Add(new LexicalComponent(lexeme, Grammar.Semicolon.ToString(), (int)Grammar.Semicolon));
    state = 0; lexeme = ""; token = "";
    break;
}
               else if (line[i] = ',') {
    lexeme += line[i];
    redd(new Lexica)
                       veature +* tine[1];
elements.Add(new LexicalComponent(lexeme, Grammar.Comma.ToString(), (int)Grammar.Comma));
state = 0; lexeme = ""; token = "";
break;
              break;
} else f(line[i] = '(') {
   lexeme ** line[i];
   elements.Add(new LexicalComponent(lexeme, Grammar.LeftParenthesis.ToString(), (int)Grammar.LeftParenthesis));
   state = 0; lexeme = ""; token = "";
   break;
}
              break;
}
else if (line[i] = ')') {
    lexeme += line[i];
    elsents.Add(new LexicalComponent(lexeme, Grammar.RightParenthesis.ToString(), (int)Grammar.RightParenthesis));
    state = 0; lexeme = ""; token = "";
    break;
}
              break;
} else f(line[i] = '{') {
    lexeme *= line[i];
    elements.Add(new LexicalComponent(lexeme, Grammar.LeftCurlyBrace.ToString(), (int)Grammar.LeftCurlyBrace));
    state = 0; lexeme = ""; token = "";
    break;
}
              break;
}
else if (line[i] = '}') {
    lexeme ** line[i];
    elsemin in [i];
    state = 0; lexeme = ""; token = "";
    break;
}
              break;
} else if (line[i] = '$') {
    lexeme =* line[i];
    elseents.Add(new LexicalComponent(lexeme, Grammar.DollarSing.ToString(), (int)Grammar.DollarSing));
    state = 0; lexeme = ""; token = "";
    break;
}
                     state = 11;
lexeme += line[i];
token = Grammar.RelationalOperator.ToString();
break;
              break;
left (line[i] = '>') {
    state = 12;
    lexeme += line[i];
    token = Grammar.RelationalOperator.ToString();
    break;
}
              break;
}
else if (line[i] = '<') {
    state = 13;
    lexeme += line[i];
    token = Grammar.RelationalOperator.ToString();
    break;
}</pre>
               }
else if (line[i] == '=') {
    state = 14;
    lexeme += lin[i];
    break;
              break;
}
else if (line[i] = '|') {
    state = 15;
    lexeme += line[i];
    token = Grammar.LogicOperator.ToString();
    break;
}
              preak;
else if (line[i] = '6') {
    state = 16;
    lexeme += line[i];
    token = Grammar.logicOperator.ToString();
    break;
              break;
else if (line[i] = '+' || line[i] = '-') {
    lexeme += line[i];
    elements.Add(new LexicalComponent(lexeme, Gram
    state = 0; lexeme = ""; token = "";
    break;
   105 else i
106 le
107 el
108 st
109 br
110 }
111 else i
112 le
113 el
114 st
115 br
116 }
117 break;
               }
else if (line[i] = '*' || line[i] = '/') {
    lexeme += line[i];
    elements.Add(new LexicalComponent(lexeme, Grammar.MultiplicationOperator.ToString(), (int)Grammar.MultiplicationOperator));
    state = 0; lexeme = ""; token = "";
    break;
}
```

Imagen 6. Case 0 del analizador léxico.

El case 1 se encarga de reconocer palabras. Aquí mismo se evalúa si el lexema corresponde a un tipo de dato, a una palabra reservada o es simplemente un identificador.

```
2 if (Char.IsDigit(line[i]) || Char.IsLetter(line[i]) || line[i] = '_') {
         state = 1;
         lexeme += line[i];
         token = Grammar.Identifier.ToString();
         if (!Char.IsDigit(line[i+1]) & !Char.IsLetter(line[i+1]) & line[i+1] ≠ '_') {
  8
             if (lexeme = "int" || lexeme = "float")
                 elements.Add(new LexicalComponent(lexeme, Grammar.DataType.ToString(), (int)Grammar.DataType));
 9
 10
             else if (lexeme = "if" || lexeme = "else" || lexeme = "while" || lexeme = "return")
 11
                 elements.Add(new LexicalComponent(lexeme, Grammar.KeyWord.ToString(), (int)Grammar.KeyWord));
 13
 14
 15
                 elements.Add(new LexicalComponent(lexeme, token, (int)Grammar.Identifier));
 16
             state = 0; lexeme = ""; token = "";
 17
 18
     }
 19
 20 break;
```

Imagen 7. Case 1 del analizador léxico. Reconocimiento de identificadores, tipos de datos y palabras reservadas.

El case 2 y 3 corresponden a números. En estos dos estados se reconocen tanto números enteros como reales.

```
1
2 case 2:
      if (Char.IsDigit(line[i])) {
          state = 2;
5
          lexeme += line[i];
          token = Grammar.Number.ToString();
 6
      }
 7
8
      else if (line[i] = '.') {
9
          state = 3;
          lexeme += line[i];
10
11
          token = Grammar.Number.ToString();
     }
12
      else {
13
14
          elements.Add(new LexicalComponent(lexeme, token, (int)Grammar.Number));
          state = 0; lexeme = ""; token = "";
15
16
17 break;
18
19 case 3:
if (char.IsDigit(line[i])) {
21
          state = 3;
          lexeme += line[i];
22
23
          token = Grammar.Number.ToString();
      }
24
25
      else {
          elements.Add(new LexicalComponent(lexeme, token, (int)Grammar.Number));
26
          state = 0; lexeme = ""; token = "";
27
       }
28
29 break;
```

Imagen 8.Case 2 y 3. Reconocimiento de números reales y enteros.

Los case del 11 al 16 se encarga de reconocer operadores relacionales, ya que los operadores relaciones se suelen componer de dos caracteres.

```
1 case 11:
       if (line[i] = '=') {
 3
           lexeme += line[i];
 4
           elements.Add(new LexicalComponent(lexeme, token, (int)Grammar.RelationalOperator));
           state = 0; lexeme = ""; token = "";
 5
 6
 7 break;
8
9 case 12:
      if (line[i] = '=') {
10
           lexeme += line[i];
11
12
           elements.Add(new LexicalComponent(lexeme, token, (int)Grammar.RelationalOperator));
13
           state = 0; lexeme = ""; token = "";
       }
14
15 break;
16
17 case 13:
      if (line[i] = '=') {
18
19
           lexeme += line[i];
           elements.Add(new LexicalComponent(lexeme, token, (int)Grammar.RelationalOperator));
20
           state = 0; lexeme = ""; token = "";
21
22
23 break;
24
25 case 14:
     if (line[i] = '=') {
26
           lexeme += line[i];
27
28
           token = Grammar.RelationalOperator.ToString();
           elements.Add(new LexicalComponent(lexeme, token, (int)Grammar.RelationalOperator));
29
           state = 0; lexeme = ""; token = "";
30
       }
31
       else {
32
33
           token = Grammar.AssignmentOperator.ToString();
           elements.Add(new LexicalComponent(lexeme, token, (int)Grammar.AssignmentOperator));
34
35
       state = 0; lexeme = ""; token = "";
36
37 break;
38
39 case 15:
      if (line[i] = '|') {
40
41
           lexeme += line[i];
           elements.Add(new LexicalComponent(lexeme, token, (int)Grammar.LogicOperator));
42
           state = 0; lexeme = ""; token = "'
43
44
45 break;
47 case 16:
      if (line[i] = '|') {
48
49
           lexeme += line[i];
50
           elements.Add(new LexicalComponent(lexeme, token, (int)Grammar.LogicOperator));
51
           state = 0; lexeme = ""; token = "";
52 break;
```

Imagen 9. Cases del 11 al 16. Reconocimiento de operadores relacionales.

El reconocimiento de los demás operadores incluidos en la gramática se hace desde el mismo case 0.

```
1 else if (line[i] = ';') {
2     lexeme += line[i];
3     elements.Add(new LexicalComponent(lexeme, Grammar.Semicolon.ToString(), (int)Grammar.Semicolon));
4     state = 0; lexeme = ""; token = "";
  10
 13 else if (line[i] = '(') {
         lexeme += line[i];
elements.Add(new LexicalComponent(lexeme, Grammar.LeftParenthesis.ToString(), (int)Grammar.LeftParenthesis));
state = 0; lexeme = ""; token = "";
 14
15
         break:
 break;
 25 else if (line[i] = '{') {
26    lexeme += line[i];
         elements.Add(new LexicalComponent(lexeme, Grammar.LeftCurlyBrace.ToString(), (int)Grammar.LeftCurlyBrace));
state = 0; lexeme = ""; token = "";
 29
 30 }
 31 else if (line[i] = '}') {
32    lexeme += line[i];
33    elsemets.Add(new LexicalComponent(lexeme, Grammar.RightCurlyBrace.ToString(), (int)Grammar.RightCurlyBrace));
34    state = 0; lexeme = ""; token = "";
 37 else if (line[i] = '$') {
         lexeme += line[i];
elements.Add(new LexicalComponent(lexeme, Grammar.DollarSing.ToString(), (int)Grammar.DollarSing));
         state = 0; lexeme = ""; token = "";
 40
         break;
 43 else if (line[i] = '!') {
         state = 11;
lexeme += line[i];
 44
 45
          token = Grammar.RelationalOperator.ToString();
         break;
 49 else if (line[i] = '>') {
         state = 12;
lexeme += line[i];
token = Grammar.RelationalOperator.ToString();
 51
 53
         break:
 55 else if (line[i] = '<') {
 56
      state = 13;
         lexeme += line[i];
token = Grammar.RelationalOperator.ToString();
 57
 59
60 }
 61 else if (line[i] = '=') {
 62
         state = 14;
lexeme += line[i];
 63
 64
         break:
 66 else if (line[i] = '|') {
         state = 15;
lexeme += line[i];
token = Grammar.LogicOperator.ToString();
 68
 70
 72 else if (line[i] = '\delta') {
         state = 16;
lexeme += line[i];
 74
          token = Grammar.LogicOperator.ToString();
 78 else if (line[i] = '+' || line[i] = '-') {
79     lexeme += line[i];
         elements.Add(new LexicalComponent(lexeme, Grammar.AdditionOperator.ToString(), (int)Grammar.AdditionOperator));
state = 0; lexeme = ""; token = "";
 81
 84 else if (line[i] = '*' || line[i] = '/') {
         e r (the[i] = * || the[i] = / ) {
lexeme += line[i];
elements.Add(new LexicalComponent(lexeme, Grammar.MultiplicationOperator.ToString(), (int)Grammar.MultiplicationOperator));
state = 0; lexeme = ""; token = "";
 85
 86
 87
         break;
 89 }
```

Imagen 10. Reconocimiento de los demás operadores en el case 0.

La última clase utilizada es la propia del Windows Form, la cual inicializa la interfaz gráfica. Además, en el constructo de la clase se hace la instancia de la lista de componentes léxicos. El método buttonLexicalAnalyzer\_Click guarda en una lista de strings el texto ingresado por el usuario, a cada línea se le agrega el operador "\$" que indica el final de una línea.

Cada una de las líneas se pasa al analizador léxico, y guarda en la lista de componentes los tokens que va reconociendo.

Finalmente, los componentes léxicos reconocidos se muestran en la tabla que está en la parte inferior de la interfaz.

Imagen 11. Código de la clase de la interfaz gráfica.

## Pruebas de Ejecución

Al ejecutar la aplicación, ingresar texto en el input señalado, y presionar el botón "Analizador Léxico", se ejecuta este análisis y se muestran los componentes léxicos del texto ingresado.



Imagen 12. Prueba de ejecución 1. Reconocimiento exitoso de componentes léxicos.

Aquí se muestra otro ejemplo de ejecución, en la que se reconocen exitosamente los componentes léxicos del texto ingresado.

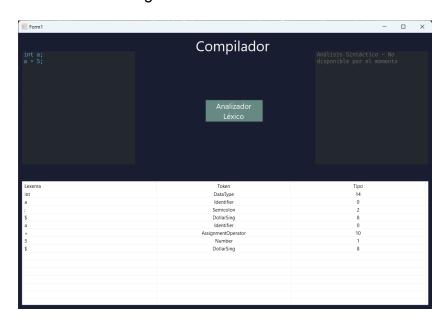


Imagen 13. Prueba de ejecución 2. Reconocimiento exitoso de componentes léxicos.

#### **CONCLUSIONES PERSONALES**

#### **Diego Alberto Martinez Zepeda:**

El trabajo llevado a cabo y presentado en este documento, permite concluir los conocimientos adquiridos y necesarios para desarrollar un analizador léxico donde se contempla la entrada en tiempo real de un usuario externo, haciendo uso de otras tecnologías y no solo de las que permiten el procesamiento, logramos desarrollar además una interfaz amigable al usuario.

#### Isaac Alberto Hernandez Chavez:

Para llevar a cabo la actividad de desarrollar un analizador léxico como parte de la aplicación de fundamentos básicos en compiladores, fue necesario el comprender como funciona la lectura de flujo en un flujo de caracteres. demostramos que este es un proceso esencial para la correcta programación de programas informáticos, es bueno comprender algo tan común para nosotros como lo es un compilador.

#### Ismael Guadalupe Avila Guerrero

Este trabajo nos permitió comprender lo complejo que es el diseño y desarrollo de compiladores. La principal complejidad fue el diseño del autómata, ya que es muy fácil perderse en la gran cantidad de estados que se llegan a tener. Una vez terminado el diseño, la implementación en código fue más sencilla. La importancia de este analizador radica en que el reconocimiento de los componentes léxicos se realice de manera correcta, ya que de ello dependen las demás etapas del proceso de compilación.

## REFERENCIAS BIBLIOGRÁFICAS