

Memoria de la Práctica de Procesadores de Lenguajes: Analizador Léxico

Diego José Abengózar Vilar, Alejandro García Castellanos,
Ignacio Javier Encinas Ramos

Grupo 82

October 30, 2019

Índice

| | | |
|----------|---|----------|
| 1 | Diseño del Analizador Léxico | 2 |
| 1.1 | Tokens | 2 |
| 1.2 | Gramática | 2 |
| 1.3 | Autómata Finito Determinista | 3 |
| 1.4 | Acciones Semánticas | 4 |
| 1.5 | Errores | 5 |
| 1.6 | Matriz de Transiciones | 5 |
| 2 | Tabla de Símbolos: Estructura e implementación | 5 |
| 3 | Anexo de Pruebas | 6 |

1 Diseño del Analizador Léxico

1.1 Tokens

<PuntoComa, - >
<CorcheteAbrir, - >
<CorcheteCerrar, - >
<ID, posTS> (Identificador)
<ENT, valor> (Dato de tipo entero)
<CAD, lex> (Dato de tipo cadena)
<ParentesisCerrar, - >
<ParentesisAbrir, - >
<SUMA, - > (Operador suma)
<MENOR, - > (Operador lógico menor)
<NOT, - > (Operador lógico de negación)
<ASIG, - > (Operador de asignación)
<ASIGOR, - > (Asignación con o lógico)
<DEC, - > (“var”)
<TipoVarENT, - > (“int”)
<TipoVarLOG, - > (“boolean”)
<TipoVarCAD, - > (“string”)
<Print, - >
<Input, - >
<Coma, - >
<Return, ->
<DECFunc, - > (“function”)
<IF, - >
<ELSE, - >

1.2 Gramática

$G(N, T, S, P)$

$S = A$

$N = \{ A, B, C, D, E, F, G, H \}$

$T = \{ del, ;, \{, \}, (,), +, <, !, =, ,, l, d, ', /, -, *, c \}$

$P:$

$A \rightarrow delA \mid ; \mid \{ \mid \} \mid (\mid) \mid + \mid < \mid ! \mid = \mid ,$

$A \rightarrow B \mid lC \mid dD \mid 'E \mid /F$

$B \rightarrow =$

$C \rightarrow lC \mid dD \mid _C \mid \lambda$

$D \rightarrow dD \mid \lambda$

$E \rightarrow cE \mid *E \mid /E \mid ' ,$

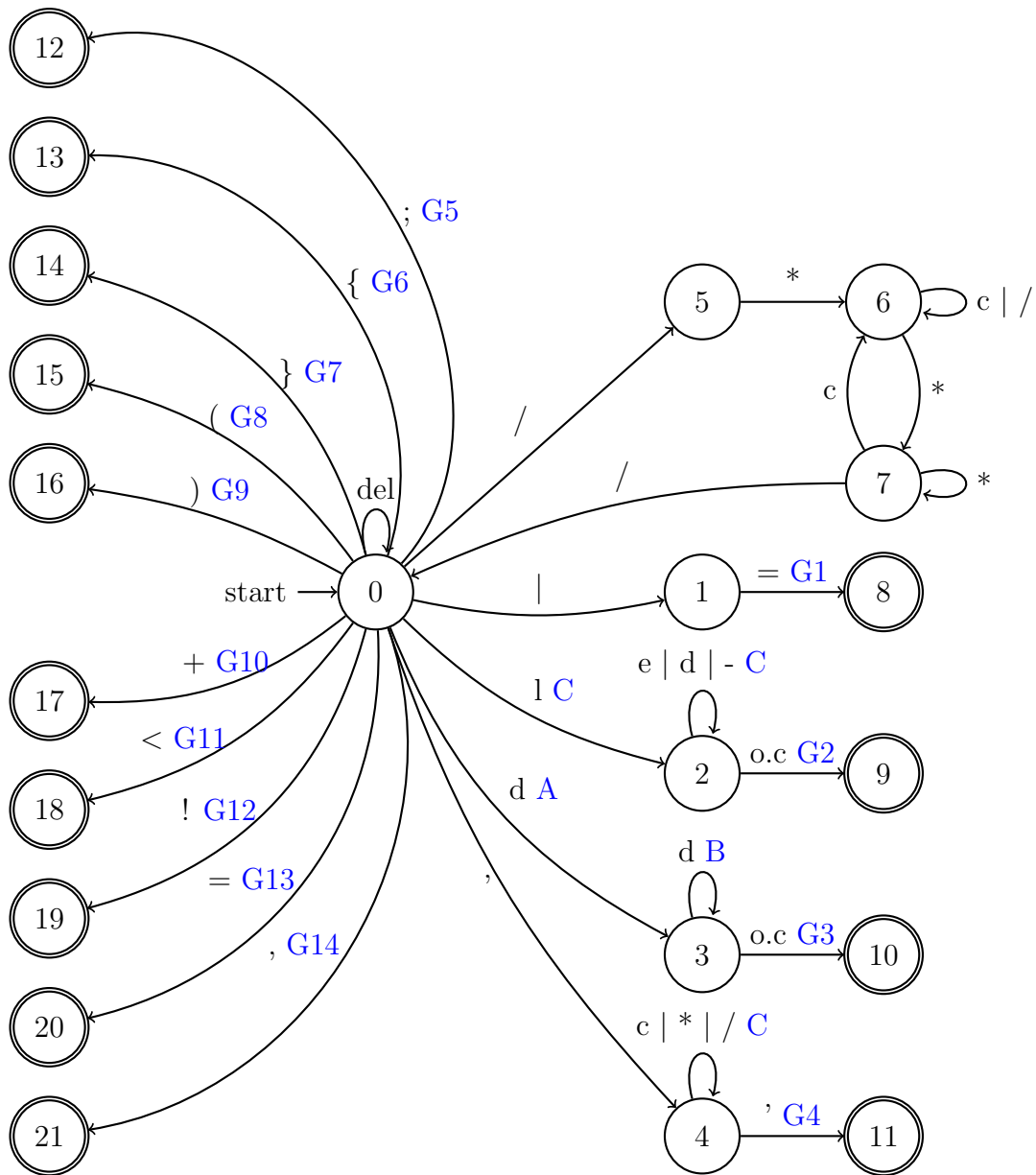
$F \rightarrow *G$

$G \rightarrow cG \mid /G \mid *H$

$H \rightarrow /A \mid cG \mid *H$

Donde, $c = T - \{*, /\}$

1.3 Autómata Finito Determinista



1.4 Acciones Semánticas

Lee \forall transicion menos o.c

C: CONCAT()

G_1 : GEN_TOKEN(ASIGOR, -)

```
 $G_2$ : if (lex  $\in$  palRes) GEN_TOKEN(palRes, -)
else if (FlagDeclUso = Decl)
    if (estaEnTSActual(lex))
        Error("Variable ya declarada")
    else
        p = INSERTAR_TS(lex)
        GEN_TOKEN(ID, p)
else
    p = BUSCA_TS(lex)
    if (p = null) p = INSERTAR_TS(lex)
    GEN_TOKEN(ID, p)
```

A: num = valor(d)

B: num = num * 10 + valor(d)

```
 $G_3$ : if (num  $\geq 2^{15}$ ) Error("Numero se sale del rango")
    else GEN_TOKEN(ENT, num)
```

G_4 : GEN_TOKEN(CAD, lex)

G_5 : GEN_TOKEN(PuntoComa, -)

G_6 : GEN_TOKEN(CorcheteAbrir, -)

G_7 : GEN_TOKEN(CorcheteCerrar, -)

G_8 : GEN_TOKEN(ParentesisAbrir, -)

G_9 : GEN_TOKEN(ParentesisCerrar, -)

G_{10} : GEN_TOKEN(SUMA, -)

G_{11} : GEN_TOKEN(MENOR, -)

G_{12} : GEN_TOKEN(NOT, -)

G_{13} : GEN_TOKEN(ASIG, -)

G_{14} : GEN_TOKEN(Coma, -)

Donde, palRes = {var, int, boolean, string, print, input, function, return, if, else}

1.5 Errores

Los errores que pueden ocurrir son errores de transiciones imprevistas, error de que un número esté fuera de rango y error de identificador ya declarado previamente (cuando el FlagDeclUso= DECL).

1.6 Matriz de Transiciones

| MT_AFD | | letra | digito | ' | / | - | carácter | * | delimitador |
|--------|----------|----------|----------|----------|----------|----------|----------|----------|-------------|
| → 0 | 1 lee | 2 C | 3A | 4 lee | 5 lee | -1 error | -1 error | -1 error | 0 lee |
| 1 | -1 error | -1 error | -1 error | -1 error | -1 error | -1 error | -1 error | -1 error | -1 error |
| 2 | 9 G2 | 2 C | 2 C | 9 G2 | 9 G2 | 2 C | 9 G2 | 9 G2 | 9 G2 |
| 3 | 10 G3 | 10 G3 | 3 B | 10 G3 | 10 G3 | 10 G3 | 10 G3 | 10 G3 | 10 G3 |
| 4 | 4 C | 4 C | 4 C | 11 G4 | 4 C | 4 C | 4 C | 4 C | 4 C |
| 5 | -1 error | -1 error | -1 error | -1 error | -1 error | -1 error | -1 error | 6 lee | -1 error |
| 6 | 6 lee | 6 lee | 6 lee | 6 lee | 6 lee | 6 lee | 6 lee | 7 lee | 6 lee |
| 7 | 6 lee | 6 lee | 6 lee | 6 lee | 0 lee | 6 lee | 6 lee | 7 lee | 6 lee |

| MT_AFD | ; | { | } | (|) | + | < | ! | = | , |
|--------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| → 0 | 12 G5 | 13 G6 | 14 G7 | 15 G8 | 16 G9 | 17 G10 | 18 G11 | 19 G12 | 20 G13 | 21 G14 |
| 1 | -1 error | -1 error | -1 error | -1 error | -1 error | -1 error | -1 error | -1 error | 8 G1 | -1 error |
| 2 | 9 G2 | 9 G2 | 9 G2 | 9 G2 | 9 G2 | 9 G2 | 9 G2 | 9 G2 | 9 G2 | 9 G2 |
| 3 | 10 G3 | 10 G3 | 10 G3 | 10 G3 | 10 G3 | 10 G3 | 10 G3 | 10 G3 | 10 G3 | 10 G3 |
| 4 | 4 C | 4 C | 4 C | 4 C | 4 C | 4 C | 4 C | 4 C | 4 C | 4 C |
| 5 | -1 error | -1 error | -1 error | -1 error | -1 error | -1 error | -1 error | -1 error | -1 error | -1 error |
| 6 | 6 lee | 6 lee | 6 lee | 6 lee | 6 lee | 6 lee | 6 lee | 6 lee | 6 lee | 6 lee |
| 7 | 6 lee | 6 lee | 6 lee | 6 lee | 6 lee | 6 lee | 6 lee | 6 lee | 6 lee | 6 lee |

2 Tabla de Símbolos: Estructura e implementación

Contiene la información de los identificadores, de los cuales se guardan los campos: lexema, tipo y desplazamiento. Para las funciones, además, se guardará el número de parámetros, su tipo, la forma de paso de parámetros y el tipo del valor de retorno.

La tabla de símbolos estará formada por dos matrices de tamaño dinámico; la primera contendrán los identificadores de ámbito global y la segunda del local. Así pues, esta segunda se creará al encontrar la declaración de una función y se borrará al acabar de ser declarada. También se utiliza un flag de declaración o uso (FlagDeclUso), un flag para saber cual es la tabla actual y dos más para el valor del desplazamiento en cada una de las tablas.

Sin embargo, en la implementación actual sólo se usa una tabla y siempre se supone que está el FlagDeclUso = Uso, pero en el caso de que no este declarada la variable se insertará en la tabla actual, ya que requerimos de la implementación del Analizador Semántico para poder saber cuando se cambia de ámbito y cuando se están declarando o usando identificadores. Así que, la acción semántica que genera los tokens de los identificadores quedaría temporalmente así:

```

G2: if (lex ∈ palRes)
      GEN_TOKEN(palRes, -)
    else if ((p:= BUSCA_TS(lex))=NULL)
      p:=INSERTAR_TS(lex)
      GEN_TOKEN(ID, p)

```

3 Anexo de Pruebas

Error 1: Número fuera de rango.

Fuente:

```

var int a;
var int b;
a = 33333;
b = a;
if (a < b) b = 1;
if (b < a) b = 8;
a = a + b;
print (a);
print (b);

```

Tokens:

```

<DEC, >
<TipoVarENT, >
<ID, 0>
<PuntoComa, >
<DEC, >
<TipoVarENT, >
<ID, 1>
<PuntoComa, >
<ID, 0>
<ASIG, >

```

Tabla Simbolos#1:

```

* LEXEMA: 'a'
* LEXEMA: 'b'

```

Error 2: Transición no prevista

Fuente:

```

var string texto; /*Comentario bueno*/
function imprime (string msg)
{
  print (msg);
}
/ Comentario malo*/
function pideTexto ()
{
  print ( 'Introduce un texto' );
  input (texto);
}

```

```

}
pideTexto();
var string textoAux;
textoAux = texto;
imprime (textoAux);

```

Tokens:

```

<DEC, >
<TipoVarCAD, >
<ID, 0>
<PuntoComa, >
<DECFunc, >
<ID, 1>
<ParentesisAbrir, >
<TipoVarCAD, >
<ID, 2>
<ParentesisCerrar, >
<CorcheteAbrir, >
<Print, >
<ParentesisAbrir, >
<ID, 2>
<ParentesisCerrar, >
<PuntoComa, >
<CorcheteCerrar, >

```

Tabla Simbolos #1:

```

* LEXEMA: 'texto'
* LEXEMA: 'imprime'
* LEXEMA: 'msg'

```

Error 3: Transición no prevista

Fuente:

```

var string texto;
function pideTexto ()
{ print ('Introduce un texto);
input (texto);
} function imprime (string msg)
{
print ('Mensaje introducido:');
print (msg);
}
pideTexto();
var string textoAux;
textoAux = texto;
imprime (textoAux);

```

Tokens:

```

<DEC, >
<TipoVarCAD, >
<ID, 0>

```

```

<PuntoComa, >
<DECFunc, >
<ID, 1>
<ParentesisAbrir, >
<ParentesisCerrar, >
<CorcheteAbrir, >
<Print, >
<ParentesisAbrir, >
<CAD, "Introduce un texto);
input (texto);
}
function imprime (string msg)
{
print (" >
<ID, 2>
<ID, 3>

```

Tabla Simbolos #1:

```

* LEXEMA: 'texto'
* LEXEMA: 'pideTexto'
* LEXEMA: 'Mensaje'
* LEXEMA: 'introducido'

```

Pruebas pasadas con éxito:

1:

Fuente:

```

var int a;
var int b;
a = 3;
b = a;
var boolean c;
c = a < b;
if (c) b = 1;
c = b < a;
if (c) b = 4;
a = a + b;
print (a);
print (b);

```

Tokens:

```

<DEC, >
<TipoVarENT, >
<ID, 0>
<PuntoComa, >
<DEC, >
<TipoVarENT, >
<ID, 1>
<PuntoComa, >
<ID, 0>

```


<ASIG, >
 <ENT, 3>
 <PuntoComa, >
 <ID, 1>
 <ASIG, >
 <ID, 0>
 <PuntoComa, >
 <DEC, >
 <TipoVarLOG, >
 <ID, 2>
 <PuntoComa, >
 <ID, 2>
 <ASIG, >
 <ID, 0>
 <MENOR, >
 <ID, 1>
 <PuntoComa, >
 <IF, >
 <ParentesisAbrir, >
 <ID, 2>
 <ParentesisCerrar, >
 <ID, 1>
 <ASIG, >
 <ENT, 1>
 <PuntoComa, >
 <ID, 2>
 <ASIG, >
 <ID, 1>
 <MENOR, >
 <ID, 0>
 <PuntoComa, >
 <IF, >
 <ParentesisAbrir, >
 <ID, 2>
 <ParentesisCerrar, >
 <ID, 1>
 <ASIG, >
 <ENT, 4>
 <PuntoComa, >
 <ID, 0>
 <ASIG, >
 <ID, 0>
 <SUMA, >
 <ID, 1>
 <PuntoComa, >
 <Print, >
 <ParentesisAbrir, >
 <ID, 0>

```

<ParentesisCerrar, >
<PuntoComa, >
<Print, >
<ParentesisAbrir, >
<ID, 1>
<ParentesisCerrar, >
<PuntoComa, >

```

Tabla Simbolos #1:

```

* LEXEMA: 'a'
* LEXEMA: 'b'
* LEXEMA: 'c'

```

2:

Fuente:

```

var int a;
var int b;
var int c;
print ( 'Introduce el primer operando' );
input (a);
print ( 'Introduce el segundo operando' );
input (b);
function int suma (int num1, int num2)
{
return num1+num2;
}
c = suma (a, b);
print (c);

```

Tokens:

```

<DEC, >
<TipoVarENT, >
<ID, 0>
<PuntoComa, >
<DEC, >
<TipoVarENT, >
<ID, 1>
<PuntoComa, >
<DEC, >
<TipoVarENT, >
<ID, 2>
<PuntoComa, >
<Print, >
<ParentesisAbrir, >
<CAD, "Introduce el primer operando">
<ParentesisCerrar, >
<PuntoComa, >
<Input, >
<ParentesisAbrir, >

```

<ID, 0>
 <ParentesisCerrar, >
 <PuntoComa, >
 <Print, >
 <ParentesisAbrir, >
 <CAD, "Introduce el segundo operando">
 <ParentesisCerrar, >
 <PuntoComa, >
 <Input, >
 <ParentesisAbrir, >
 <ID, 1>
 <ParentesisCerrar, >
 <PuntoComa, >
 <DECFunc, >
 <TipoVarENT, >
 <ID, 3>
 <ParentesisAbrir, >
 <TipoVarENT, >
 <ID, 4>
 <Coma, >
 <TipoVarENT, >
 <ID, 5>
 <ParentesisCerrar, >
 <CorcheteAbrir, >
 <Return, >
 <ID, 4>
 <SUMA, >
 <ID, 5>
 <PuntoComa, >
 <CorcheteCerrar, >
 <ID, 2>
 <ASIG, >
 <ID, 3>
 <ParentesisAbrir, >
 <ID, 0>
 <Coma, >
 <ID, 1>
 <ParentesisCerrar, >
 <PuntoComa, >
 <Print, >
 <ParentesisAbrir, >
 <ID, 2>
 <ParentesisCerrar, >
 <PuntoComa, >

Tabla Simbolos #1:

- * LEXEMA: 'a'
- * LEXEMA: 'b'
- * LEXEMA: 'c'

* LEXEMA: 'suma'
* LEXEMA: 'num1'
* LEXEMA: 'num2'

3:

Fuente:

```
var int a;  
var int b;  
a = 3;  
b = a;  
if (a < b) b = 1;  
a = a + b;  
print (a);  
print (b);
```

Tokens:

```
<DEC, >  
<TipoVarENT, >  
<ID, 0>  
<PuntoComa, >  
<DEC, >  
<TipoVarENT, >  
<ID, 1>  
<PuntoComa, >  
<ID, 0>  
<ASIG, >  
<ENT, 3>  
<PuntoComa, >  
<ID, 1>  
<ASIG, >  
<ID, 0>  
<PuntoComa, >  
<IF, >  
<ParentesisAbrir, >  
<ID, 0>  
<MENOR, >  
<ID, 1>  
<ParentesisCerrar, >  
<ID, 1>  
<ASIG, >  
<ENT, 1>  
<PuntoComa, >  
<ID, 0>  
<ASIG, >  
<ID, 0>  
<SUMA, >  
<ID, 1>  
<PuntoComa, >
```

<Print, >
<ParentesisAbrir, >
<ID, 0>
<ParentesisCerrar, >
<PuntoComa, >
<Print, >
<ParentesisAbrir, >
<ID, 1>
<ParentesisCerrar, >
<PuntoComa, >

Tabla Simbolos #1:

* LEXEMA: 'a'

* LEXEMA: 'b'