



CARRERA INGENIERÍA DE SOFTWARE

Docente:

Juan Vilca Castro

Curso:

Compiladores

Alumnos:

Butrón Soria, Jarol

Milón Guzman, Gianella

Mendoza Melgar, Rodrigo

Loza Peralta, Christian

Tema:

**DOCUMENTACIÓN DE UN ANALIZADOR LÉXICO Y SINTÁCTICO,
PARA LA CONSTRUCCIÓN DE UN COMPILADOR**

Saturday 22nd October, 2016

Resumen

El presente documento detalla la información para la construcción de un analizador léxico y sintáctico que formará parte del compilador para un propio lenguaje definido, construido bajo la plataforma de Java, asimismo se especifica la lista de palabras reservadas, tokens, matrices, operadores, delimitadores, identificadores, literales, constantes y definición formal del autómeta.

1. Lista de *Tokens*

1.1. Palabras reservadas

- **programa:** bloque de partida para la ejecución del programa.
- **entero:** Tipo de dato entero de 32 Bits (4 bytes), conformado por números positivos.
- **si:** Estructura condicional o instrucción de control simple, evalúa una expresión dada.
- **osino:** Instrucción de control doble, permite especificar una acción cuando la condición es verdadera y distinta cuando la condición es falsa.
- **mientras:** Instrucción de repetición, evalúa una expresión, en caso de devolver verdadero iniciará un ciclo (bucle) hasta que la condición sea verdadera.
- **leerTeclado:** Permite la lectura de datos mediante teclado.
- **escribirPantalla:** Muestra la salida de datos en pantalla.

1.2. Operadores

1.2.1. Operadores unarios

- **Operador "-" (menor):** Indica la negación de una variable o una expresión.
- **Operador "+" (mayor):** Indica un valor positivo de una variable o una expresión.

1.2.2. Operadores binarios

- **Operador " + " (suma) :** Operador de adición.
- **Operador " - " (resta):** Operador de resta.
- **Operador " * " (multiplicación):** Operador de multiplicación.
- **Operador " / " (resta):** Operador de división.
- **Operador " % " (módulo):** Operador de residuo de división.
- **Operador " < " (menor):** Operador de relación menor.
- **Operador " > " (mayor):** Operador de relación mayor.
- **Operador " <= " (menor e igual):** Operador de relación menor e igual.
- **Operador " >= " (Mayor igual):** Operador de relación mayor e igual.
- **Operador " = " (igual):** Operador de asignación.
- **Operador " == " (exactamente igual):** Operador de relación de igualdad de valor.
- **Operador " != " (diferencia):** Operador de relación diferente entre valores de datos.
- **Operador " && " (OR):** Operador *OR* lógico.
- **Operador " && " (AND):** Operador *AND* lógico.

2.2. Representación del autómata finito determinista

estadoActual	estadoSiguiente	caracteresAceptados
0	8	ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz_\$
0	7	0123456789
0	5	{};:
0	4	<>!=
0	3	+-%*
0	9	&
0	6	
0	13	/
0	1	"
1	1	/ &+-%*<>!={};:_\$0123456789ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz
1	2	"
4	3	=
6	3	
7	7	0123456789
8	8	0123456789ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz
9	3	&
10	10	/ &+-%*<>!={};:_\$0123456789ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz
10	11	*
11	12	/
11	10	* &+-%*<>!={};:_\$0123456789ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz
13	14	/
13	10	*
14	14	/ &+-%*<>!={};:_\$0123456789ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz
14	15	

Figura. 2: Representación del autómata finito determinista

3. Analizador Sintáctico

3.1. Gramática

1. $R_0 : < P > \rightarrow \text{Programa}() < BS >$
2. $R_1 : < BS > \rightarrow < S^* >$
3. $R_1 : < BS > \rightarrow < S^* >$
4. $R_2 : < S^* > \rightarrow < S > < S^* >$
5. $R_3 : < S^* > \rightarrow$
 λ
6. $R_4 : < S > \rightarrow < SD >$
7. $R_5 : < S > \rightarrow < SA >$
8. $R_6 : < S > \rightarrow < SC >$
9. $R_7 : < S > \rightarrow < SM >$
10. $R_8 : < S > \rightarrow < SL >$
11. $R_9 : < S > \rightarrow < SE >$
12. $R_{10} : < S > \rightarrow < SHM >$
13. $R_{11} : < SD > \rightarrow < TD > < VA > < RD^* >;$

14. $R12 : \langle RD* \rangle \rightarrow, \langle VA \rangle \langle RD* \rangle$
15. $R13 : \langle RD* \rangle \rightarrow \lambda$
16. $R14 : \langle VA \rangle \rightarrow IDENTIFICADOR \langle A \rangle$
17. $R15 : \langle A \rangle \rightarrow = \langle EM \rangle$
18. $R16 : \langle A \rangle \rightarrow \lambda$
19. $R17 : \langle TD \rangle \rightarrow CONST_N UMERICA$
20. $R18 : \langle SA \rangle \rightarrow IDENTIFICADOR = \langle EM \rangle;$
21. $R19 : \langle SM \rangle \rightarrow Mientras(\langle EM \rangle) \langle SubB \rangle$
22. $R20 : \langle SubB \rangle \rightarrow \langle S \rangle$
23. $R21 : \langle SubB \rangle \rightarrow \langle BS \rangle$
24. $R22 : \langle SC \rangle \rightarrow Si(\langle EM \rangle) \langle SubB \rangle \langle ELSE \rangle$
25. $R23 : \langle ELSE \rangle \rightarrow Osino \langle SubB \rangle$
26. $R24 : \langle ELSE \rangle \rightarrow \lambda$
27. $R25 : \langle SHM \rangle \rightarrow Hacer \langle subB \rangle Mientras(\langle EM \rangle);$
28. $R26 : \langle EM \rangle \rightarrow \langle T \rangle \langle T* \rangle$
29. $R27 : \langle T* \rangle \rightarrow \lambda$
30. $R28 : \langle T* \rangle \rightarrow \langle OP1 \rangle \langle T \rangle \langle T* \rangle$
31. $R29 : \langle OP1 \rangle \rightarrow +$
32. $R30 : \langle OP1 \rangle \rightarrow -$
33. $R31 : \langle T \rangle \rightarrow \langle F \rangle \langle F* \rangle$
34. $R32 : \langle F* \rangle \rightarrow \langle OP2 \rangle \langle F \rangle \langle F* \rangle$
35. $R33 : \langle F* \rangle \rightarrow \lambda$
36. $R34 : \langle F \rangle \rightarrow (\langle EM \rangle)$
37. $R35 : \langle F \rangle \rightarrow CONST_N UMERICA$
38. $R36 : \langle F \rangle \rightarrow IDENTIFICADOR$
39. $R37 : \langle OP2 \rangle \rightarrow /$
40. $R38 : \langle OP2 \rangle \rightarrow \%$
41. $R39 : \langle OP2 \rangle \rightarrow *$
42. $R40 : \langle SL \rangle \rightarrow LeerTeclado(IDENTIFICADOR)$
43. $R41 : \langle SE \rangle \rightarrow EscribirPantalla(\langle E \rangle);$
44. $R42 : \langle E \rangle \rightarrow \langle EM \rangle$
45. $R43 : \langle E \rangle \rightarrow IDENTIFICADOR$
46. $R44 : \langle E \rangle \rightarrow LITERAL_CADENA$
47. $R45 : \langle E \rangle \rightarrow CONST_N UMERICA$
48. $R46 : \langle E \rangle \rightarrow \lambda$

Terminología

1. Terminal

- Programa
- (
-)
- {
- }
- ;
- ,
- =
- Si
- Osino
- Hacer
- Mientras
- +
- −
- /
- *
- %
- LeerTeclado
- EscribirPantalla
- IDENTIFICADOR
- CONST_NUMERICA
- LITERAL_CADENA

2. No Terminales

- Programa
- $\langle P \rangle$ Programa
- $\langle BS \rangle$ (bloquedesentencias)
- $\langle S^* \rangle$ (Listadesentencias)
- $\langle SD \rangle$ (Sentenciadedeclaración)
- $\langle TD \rangle$ (tipodedato)
- $\langle VA \rangle$ (variableasignada)
- $\langle RD^* \rangle$ (restodedatos)
- $\langle A \rangle$ (asignacin)
- $\langle EM \rangle$ (expresinmatemática)
- $\langle SA \rangle$ (sentenciadeasignación)
- $\langle SC \rangle$ (sentenciadecondicional)
- $\langle SM \rangle$ (sentenciademientras)
- $\langle SL \rangle$ (sentenciadelectura)
- $\langle SE \rangle$ (sentenciadeescritura)
- $\langle SubB \rangle$ (subbloque)
- $\langle ELSE \rangle$ (condicionalalternativa)
- $\langle SHM \rangle$ (sentenciahacermientras)

- $\langle T \rangle$ (Término)
- $\langle T^* \rangle$ (ListadeTérmino)
- $\langle OP1 \rangle$ (Operadoresnivel1)
- $\langle OP2 \rangle$ (Operadoresnivel2)
- $\langle F \rangle$ (Factor)
- $\langle F^* \rangle$ (ListadeFactores)
- $\langle E \rangle$ (escribir)

3. lamdba

- lambda (λ) o vacío.

3.2. Representación del analizador sintáctico

noTerminal	derivacion		
	tipoToken	lexema	valor
		Programa	
valor <P>		(
)	
			<BS>
	tipoToken	lexema	valor
		{	
valor <BS>			<S*>
		}	
	valor		
valor <S*>	<S>		
	<S*>		
valor <S*>			
valor <S>	valor		
	<SD>		
valor <S>	valor		
	<SA>		
valor <S>	valor		
	<SC>		
valor <S>	valor		
	<SM>		
valor <S>	valor		
	<SL>		

Figura. 3: representación del analizador sintáctico 01

valor <S>	valor		
	<SE>		
valor <S>	valor		
	<SHM>		
	valor	tipoToken	lexema
	<TD>		
valor <SD>	<VA>		
	<RD*>		
			;
	tipoToken	lexema	valor
valor <RD*>		,	
			<VA>
			<RD*>
valor <RD*>			
	tipoToken	lexema	valor
valor <VA>	IDENTIFICADOR		
			<A>
	tipoToken	lexema	valor
valor <A>		=	
			
valor <A>			
valor <TD>	tipoToken	lexema	
		entero	

Figura. 4: representación del analizador sintáctico 02

<div></div>		tipoToken	lexema	valor
		IDENTIFICADOR		
			=	
				
valor	<SA>			
<div></div>		tipoToken	lexema	valor
			Mientras	
			(
				
)	
				<SubB>
valor	<SM>			
<div></div>		valor		
		<S>		
valor	<SubB>			
<div></div>		valor		
		<BS>		
<div></div>		tipoToken	lexema	valor
			Si	
			(
				
)	
				<SubB>
valor	<SC>			
<div></div>				<ELSE>
		valor		
		Osino		
				<SubB>
valor	<ELSE>			

Figura. 5: representación del analizador sintáctico 03

valor	<ELSE>			
valor	<SHM>	tipoToken	lexema	valor
			Hacer	
				<SubB>
			Mientras	
		(
				
)		
			;	
		valor		
valor		<T>		
		<T*>		
		valor		
		<op1>		
valor	<T*>	<T>		
		<T*>		
valor	<T*>			
		tipoToken	lexema	
valor	<op1>		+	
		tipoToken	lexema	
valor	<op1>		-	
		valor		
valor	<T>	<F>		
		<F*>		

Figura. 6: representación del analizador sintáctico 04

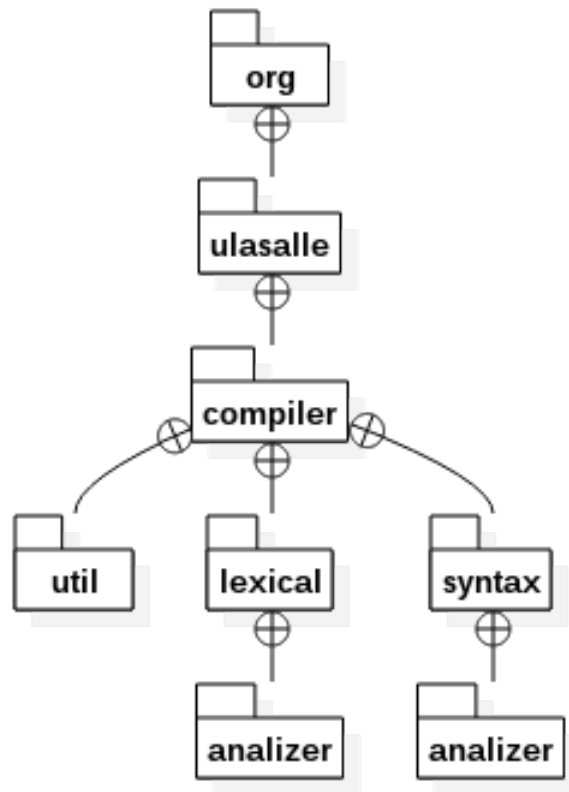
		valor		
		<op2>		
valor	<F*>	<F>		
		<F*>		
valor		<F*>		
		tipoToken	lexema	valor
			(
valor	<F>			
)	
		tipoToken		lexema
valor	<F>	CONST_NUMERICA		
		tipoToken		lexema
valor	<F>	IDENTIFICADOR		
		tipoToken		lexema
valor	<op2>			*
		tipoToken		lexema
valor	<op2>			/
		tipoToken		lexema
valor	<op2>			%
		tipoToken		lexema
				LeerTeclado
valor	<SL>			(
		IDENTIFICADOR		
)

Figura. 7: representación del analizador sintáctico 05

		tipoToken	lexema	valor
			EscribirPantalla	
valor	<SE>		(
				<E>
)	
			;	
valor		valor		
				
		tipoToken		lexema
valor	<E>	IDENTIFICADOR		
		tipoToken		lexema
valor	<E>	LITERAL_CADENA		
		tipoToken		lexema
valor	<E>	CONST_NUMERICA		

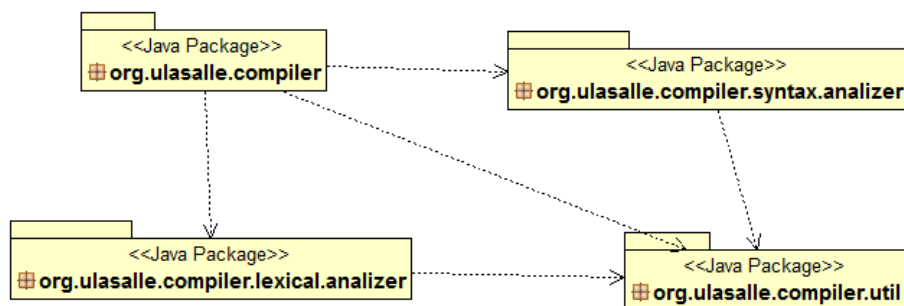
Figura. 8: representación del analizador sintáctico 06

Figura. 9: Jerarquía de los paquetes del programa mediante un Diagrama de paquetes.



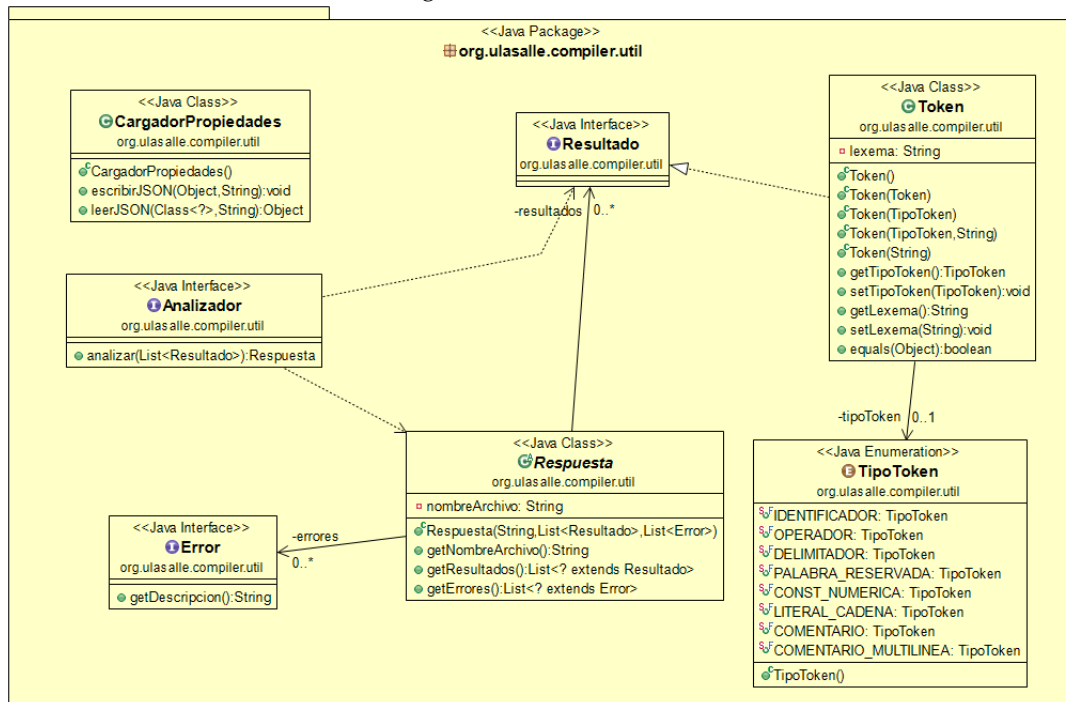
Fuente: Propia.

Figura. 10: Diagrama de paquetes.



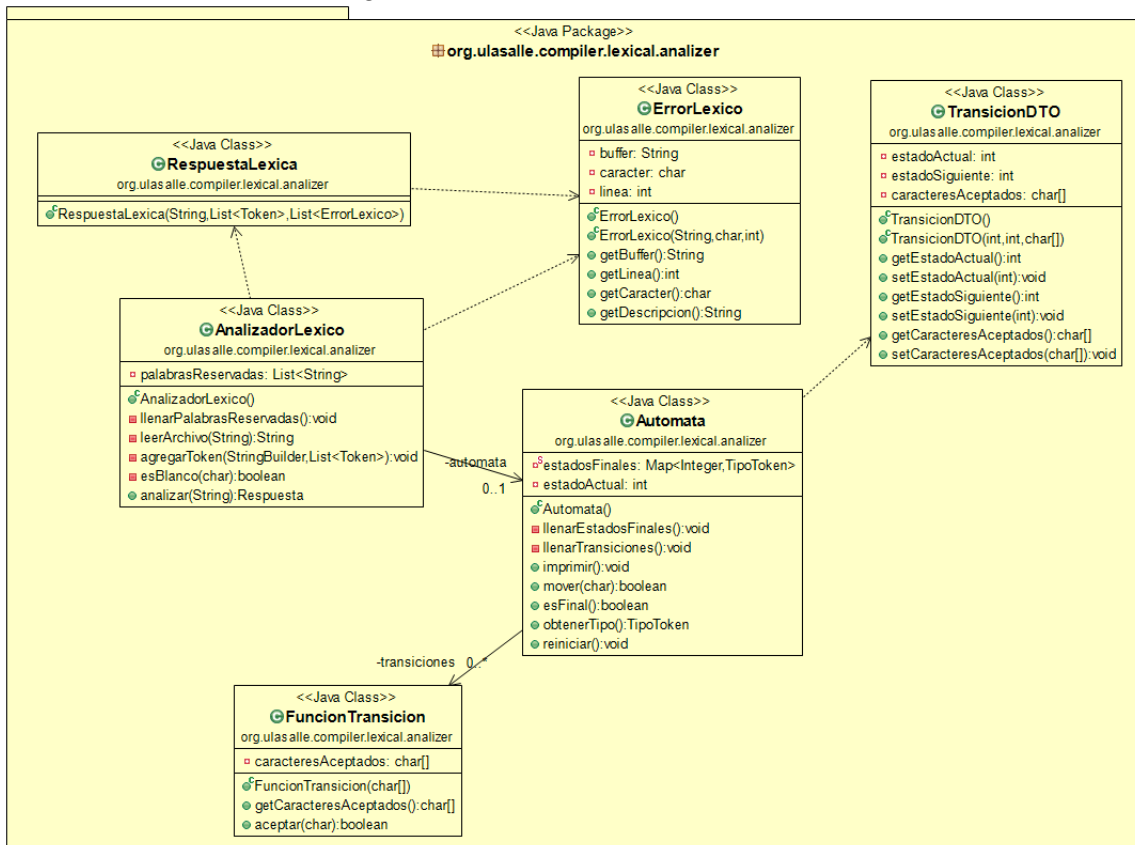
Fuente: Propia.

Figura. 11: Clases utiles.



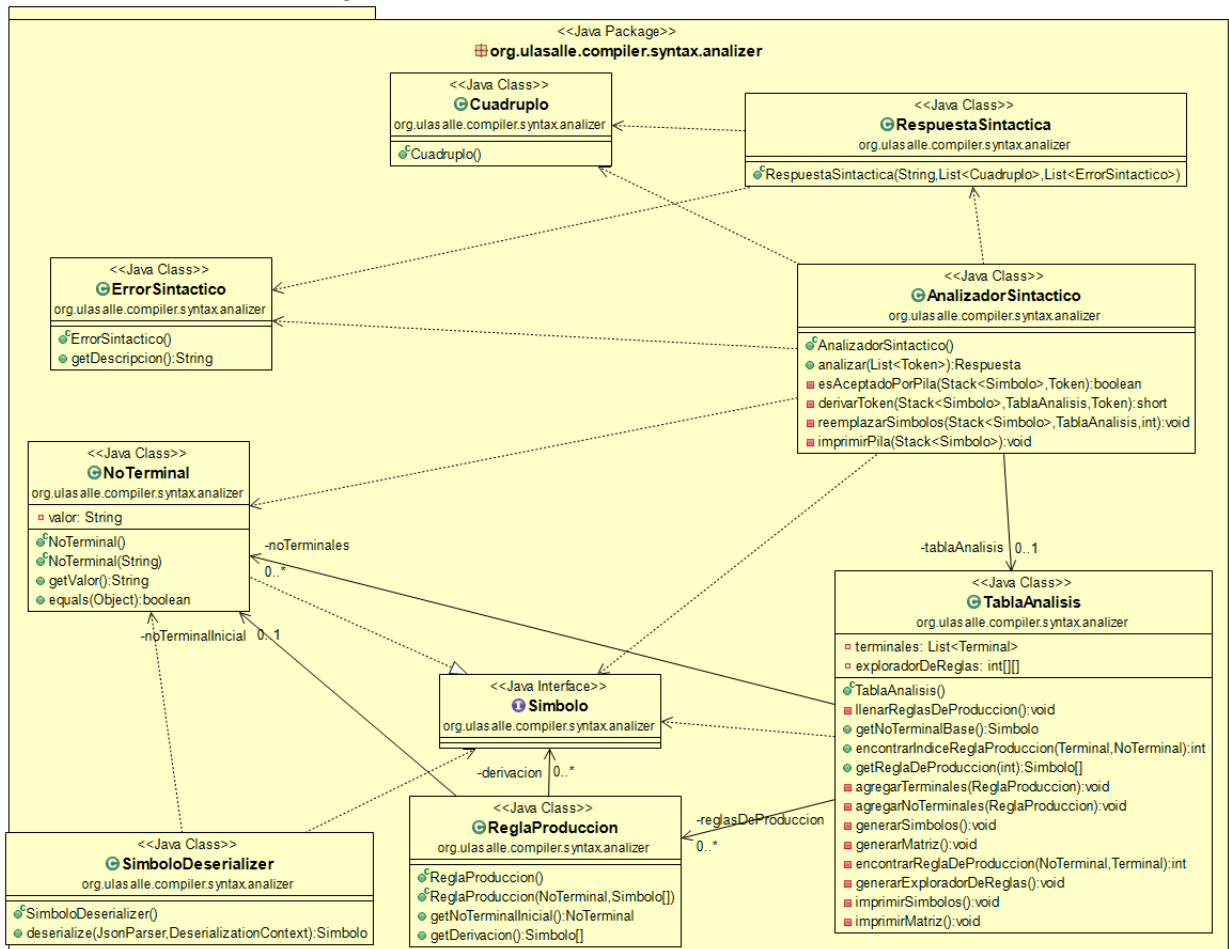
Fuente: Propia.

Figura. 12: Clases de analizador léxico.



Fuente: Propia.

Figura. 13: Clases de analizador sintactico.



Fuente: Propia.