

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA.
CENTRO UNIVERSITARIO DE OCCIDENTE.
DIVISION DE CIENCIAS DE LA INGENIERÍA.
INGENIERO OLIVER ERNESTO SIERRA PAC.
LEGUAJES FORMALES Y DE PROGRAMACION “A”.



PROYECTO FINAL.

ESTUARDO DAVID BARRENO NIMATUJ.
CARNÉ: 201830233.

ANÁLISIS LÉXICO.

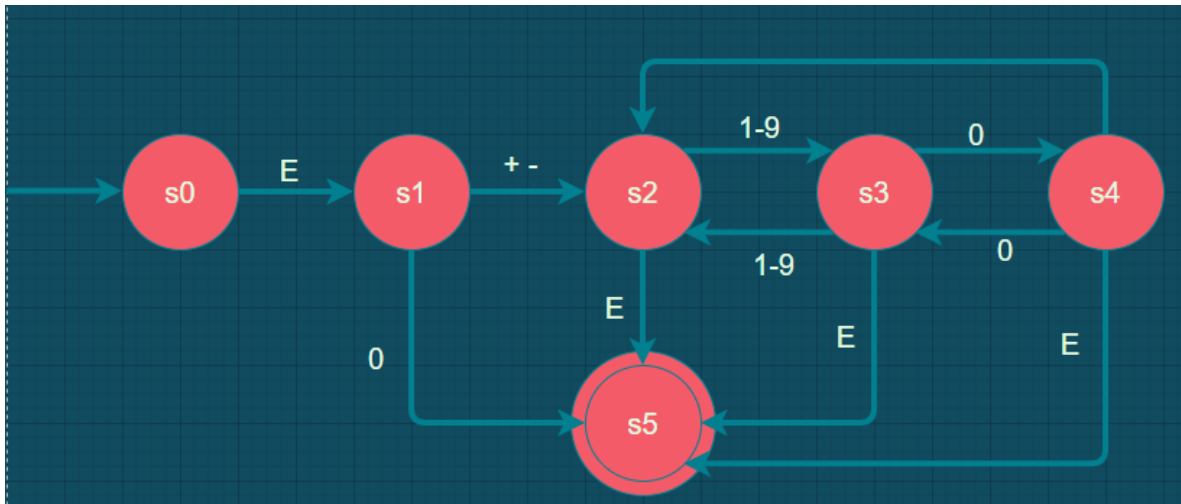
El Análisis Léxico es la primera fase de un compilador y consiste en un programa que recibe como entrada el código fuente de otro programa como la secuencia de caracteres y produce una salida compuesta de tokens o componentes léxicos o símbolos. Los tokens sirven para una posterior etapa del proceso de traducción, siendo la entrada para el Análisis Sintáctico. Su principal función consiste en leer los caracteres de entrada y elaborar como salida una secuencia de componentes léxicos que utiliza el analizador sintáctico para hacer el análisis.

Número Entero.

Número entero, puede ser negativo (no se aceptan enteros con ceros sin valor al inicio).

Alfabeto: $\{\{0\},\{+-\},\{1-9\}\} = \{0,+-,1-9\}$

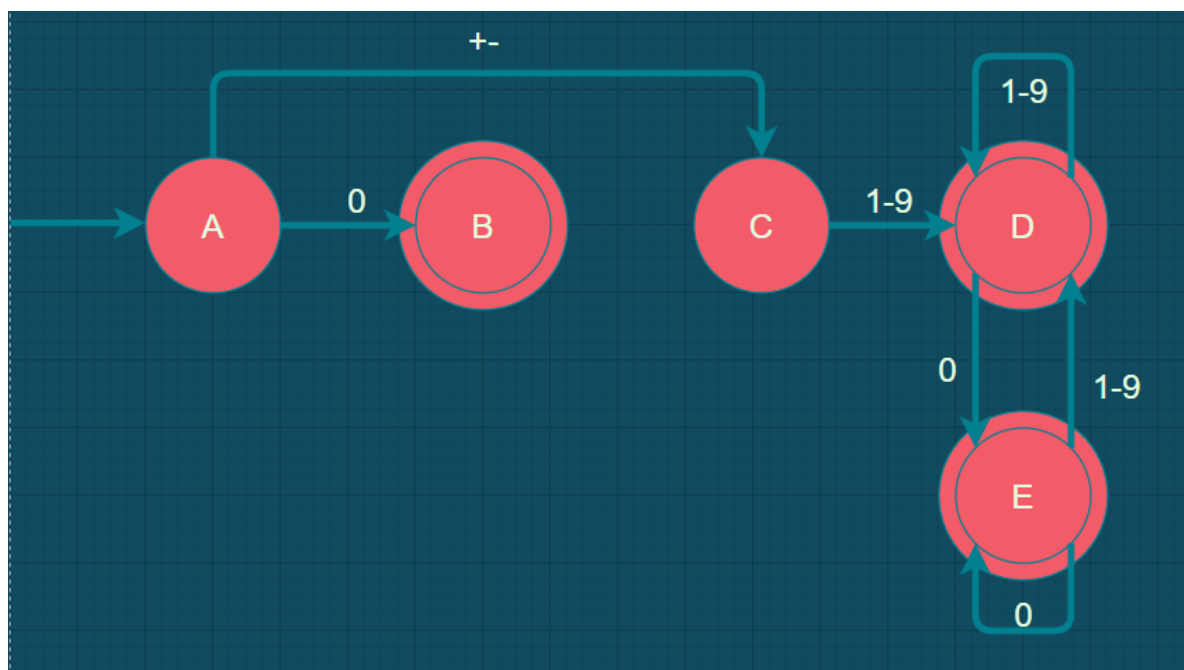
Expresión regular: $(0 | ((+|-)((1-9)|((1-9)*(0)*))))$



FT	S	0	+	-	1-9.
s1	{s1}=A	(A,0)=s5=B	(A,+)=s2=C	(A,-)=s2=C	(A,1-9)={}
s2,s5	{}=B	(B,0)={}	(B,+)=s2=C	(B,-)=s2=C	(B,1-9)={}
s2,s5	{}=C	(C,0)={}	(C,+)=s2=C	(C,-)=s2=C	(C,1-9)=s3,s5=D
s3	{}=D	(D,0)=s4,s5=E	(D,+)=s2=C	(D,-)=s2=C	(D,1-9)=s2,s5=D
s4	{}=E	(E,0)=s5=E	(E,+)=s2=C	(E,-)=s2=C	(E,1-9)=s5=D

Reducido:

FT	0	+	-	1 - 9.
A	B	C	C	
B				
C				D
D	E			D
E	E			D

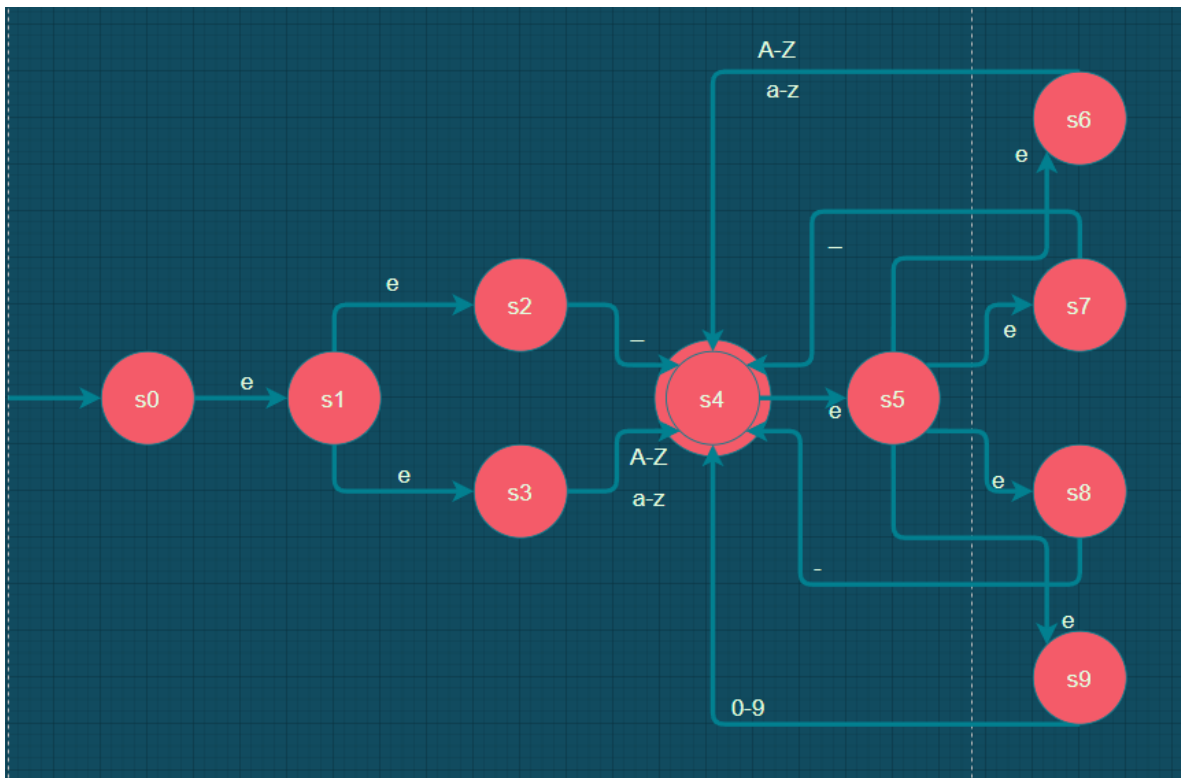


Identificador.

inicia con guiones bajos o letras seguidas de más letras o dígitos o guiones medios (-) o guiones bajos (_), puede tener longitud de un carácter o más

Alfabeto: $\{\{_ \}, \{A-Z\}, \{a-z\}, \{- \}, \{0-9\}\} = \{_, A-Z, a-z, 0-9, - \}$

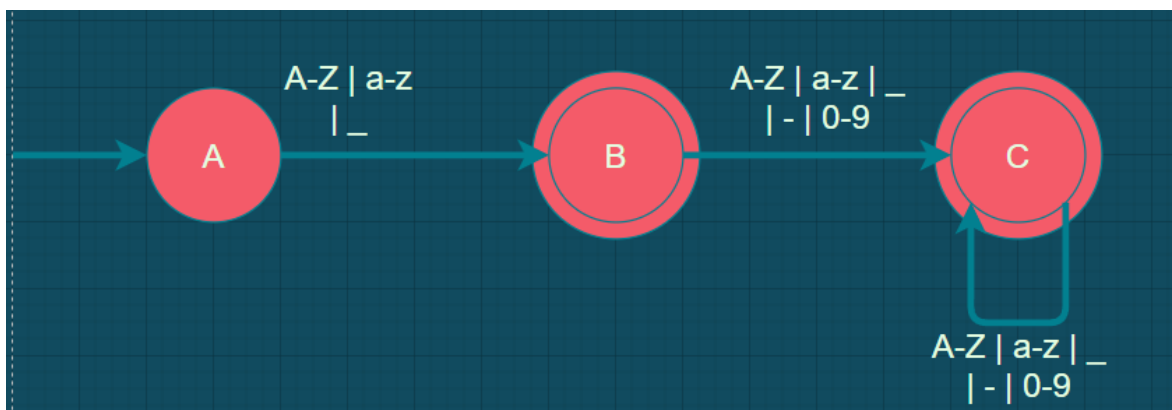
Expresión regular: $((_)|(A-Z)|(a-z))((A-Z)|(a-z)|(-)|(_)|(0-9))^*$



FT	S	A-Z	a-z	-	-	0-9
s0	{s1,s2,s3}=A	(A,A-Z)=s4=B	(A,a-z)=s4=B	(A,_) =s4=B	(A,-)={}	(A,0-9)={}
s4	{s5,s6,s7,s8,s9}=B	(B,A-Z)=s4=c	(B,a-z)=s4=c	(B,_) =s4=c	(B,-)=s4=c	(B,0-9)=s4=c
s4	{s5,s6,s7,s8,s9}=C	(C,A-Z)=s4=c	(C,a-z)=s4=c	(C,_) =s4=c	(C,-)=s4=c	(C,0-9)=s4=c

Reducido:

FT	A-Z	a-z.	_	-	0-9
A	B	B	B		
B	C	C	C	C	C
C	C	C	C	C	C

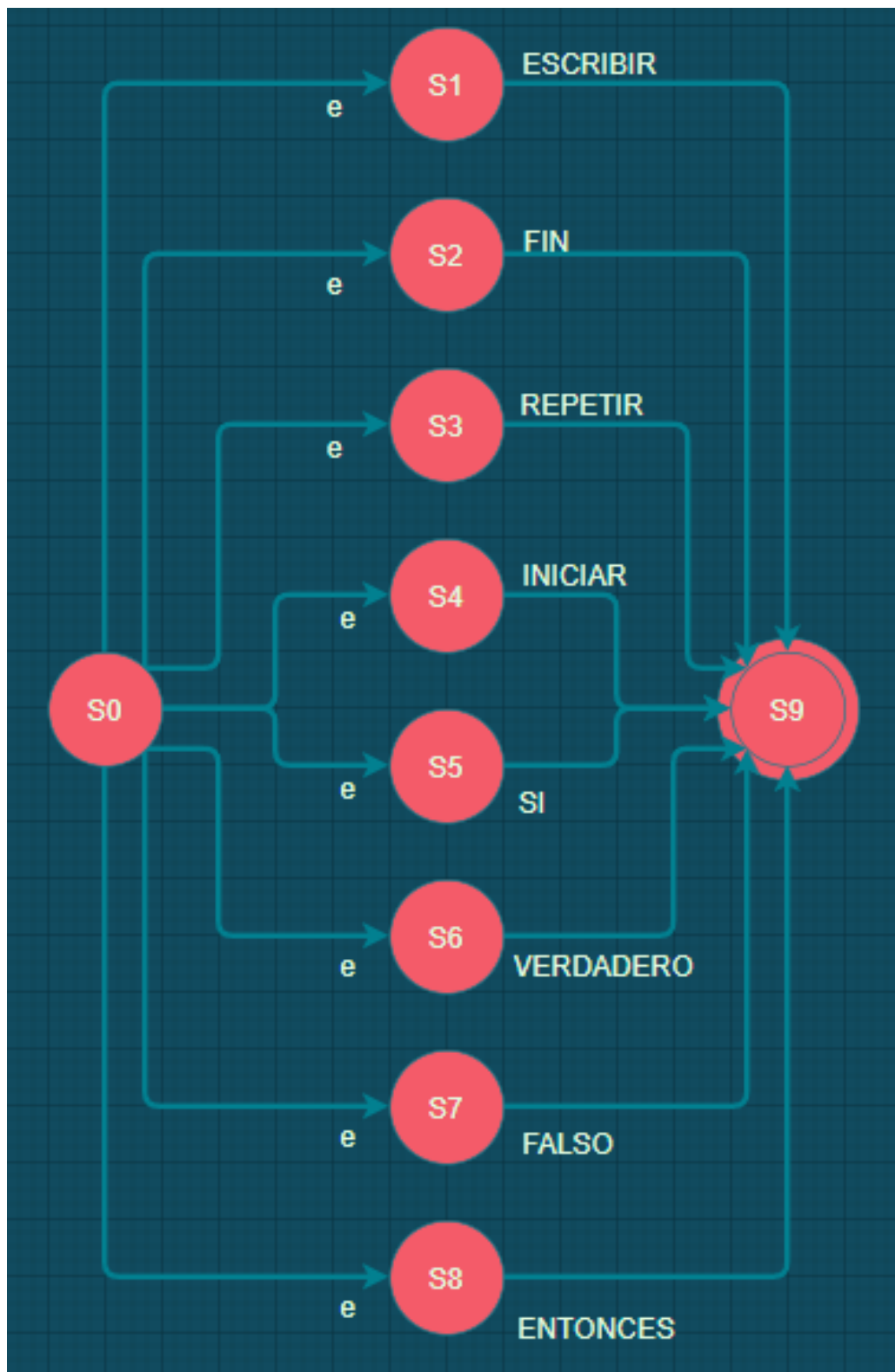


Palabras Reservadas.

Palabras reservadas: ESCRIBIR, FIN, REPETIR, INICIAR, SI, VERDADERO, FALSO, ENTONCES.

Alfabeto: $\{\{ \text{ESCRIBIR} \}, \{ \text{FIN} \}, \{ \text{REPETIR} \}, \{ \text{INICIAR} \}, \{ \text{SI} \}, \{ \text{VERDADERO} \}, \{ \text{FALSO} \}, \{ \text{ENTONCES} \}\} = \{ \text{ESCRIBIR}, \text{FIN}, \text{REPETIR}, \text{INICIAR}, \text{SI}, \text{VERDADERO}, \text{FALSO}, \text{ENTONCES} \}$

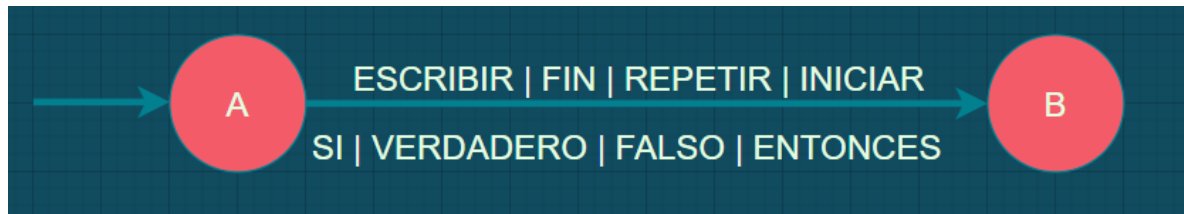
Expresión regular: $((\text{ESCRIBIR}) \mid (\text{FIN}) \mid (\text{REPETIR}) \mid (\text{INICIAR}) \mid (\text{SI}) \mid (\text{VERDADERO}) \mid (\text{FALSO}) \mid (\text{ENTONCES}))^*$



FT	S	PalabrasR.
s0	{s1,s2,s3,s4,s5,s6,s7,s8}=A	(A,PalabrasR.)= s9 =B

Reducido:

FT	PalabrasR.
A	B
B	

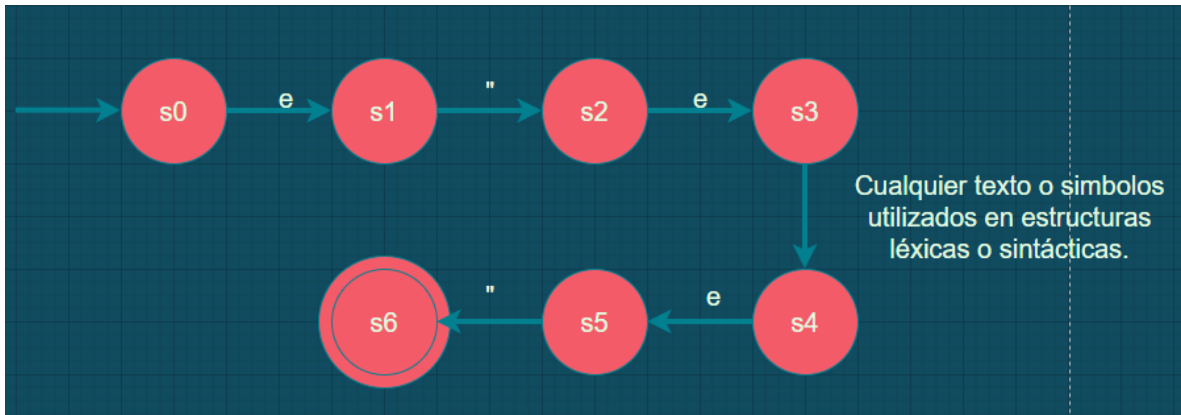


Literal.

Cualquier conjunto de caracteres envuelto en comillas (no se aceptan saltos de línea en literales).

Alfabeto: $\{\{“\}, \{\text{Cualquier texto o símbolos utilizados en estructuras léxicas o sintácticas}\}, \{”\}\} = \{\text{“Cualquier texto o símbolos utilizados en estructuras léxicas o sintácticas”}\}$

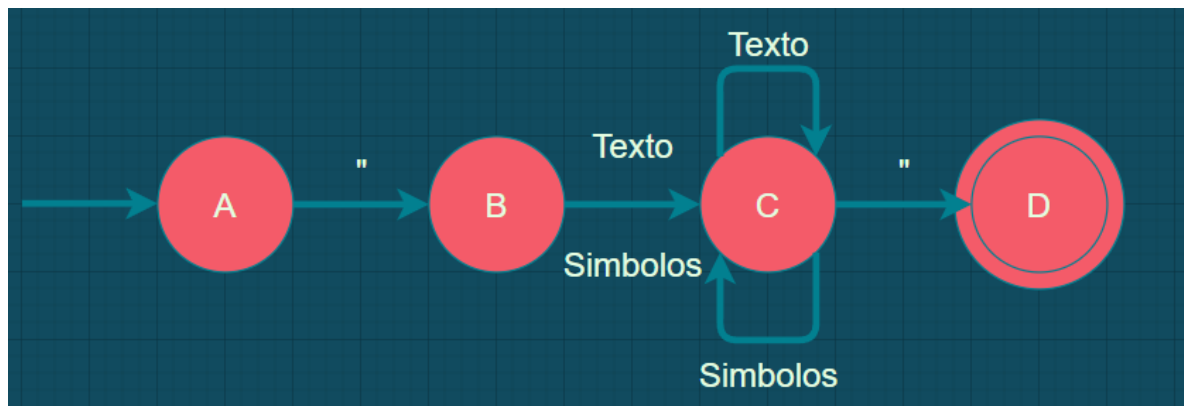
Expresión regular: $((“)(\text{Cualquier texto o símbolos utilizados en estructuras léxicas o sintácticas})* (“))$



FT	s	"	Texto/símbolos	".
s0	{s1}=A	(A,")=s2=B	(A,T/S)={}	(A,")={}
s2	{s3}=B	(B,")={}	(B,T/S)=s4=C	(B,")={}
s4	{s5}=C	(C,")={}	(C,T/S)=C	(C,")=s6=D
s6	{}=D	(D,")={}	(D,T/S)={}	(D,")={}

Reducido:

FT	"	Texto/simbolos	"	".
A	B			
B		C		
C		C		D
D				

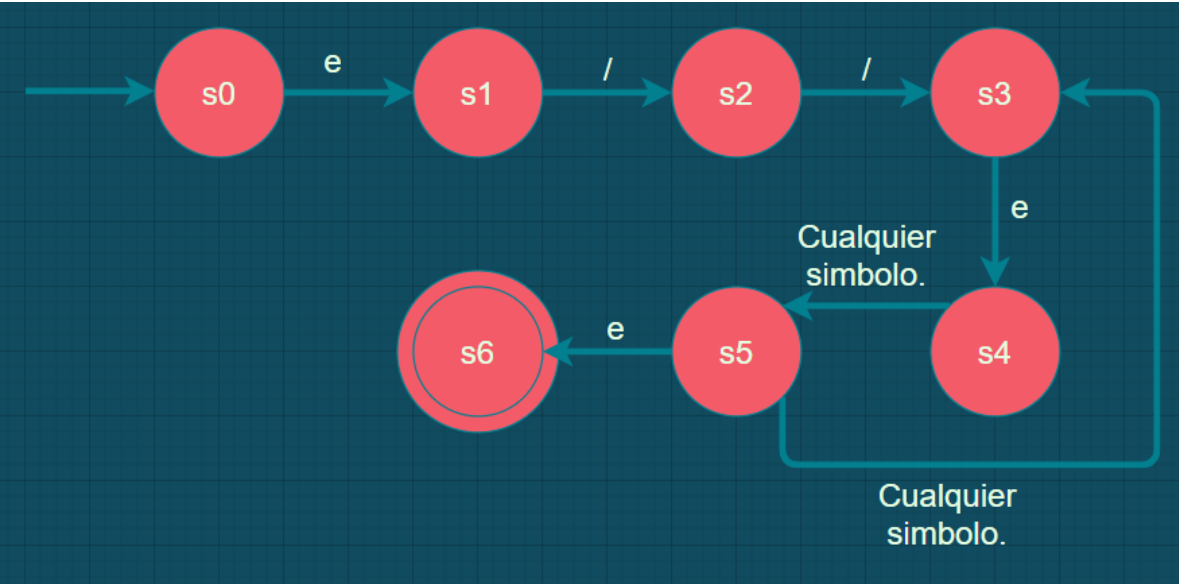


Comentarios.

Inician con dos diagonales seguidos de cualquier símbolo, el fin de línea indica el final del comentario.

Alfabeto: { {//}, {Cualquier símbolo} } = { //Cualquier simbolo }

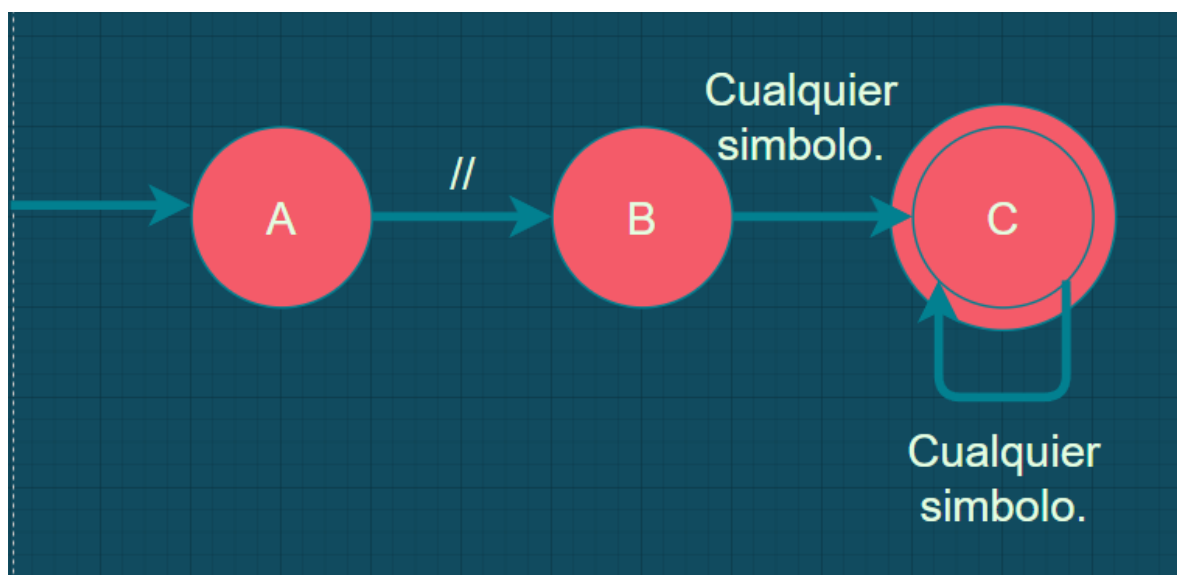
Expresión Regular: ((/)(/)(Cualquier simbolo)*)



FT	s	//	Texto/simbolos
s0	{s1}=A	(A,//)=B	(A,T/S)={}
s1	{s2,s3,s4}=B	(B,//)={}	(B,T/S)=s5,s6=C
s5,s6	{s3,s4}=C	(C,//)={}	(C,T/S)=s5,s6=C

Reducido:

FT		//		Texto/simbolos
A		B		
B				C
C				C

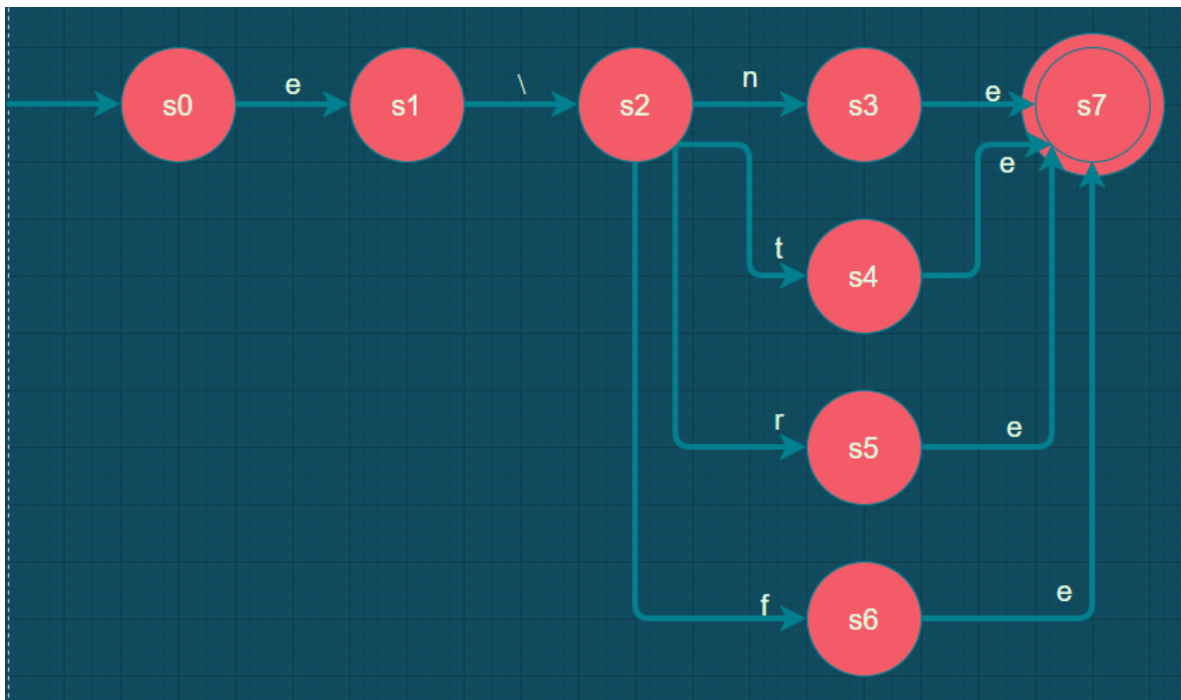


Caracteres Especiales.

Como espacio, salto de línea (`\n`), tabulaciones (`\t`), retorno de carro (`\r`), salto de página (`\f`) deben ser aceptados por el lenguaje pero ignorados al momento de análisis ya que indican separación de palabras.

Alfabeto: $\{\{\backslash n\}, \{\backslash t\}, \{\backslash r\}, \{\backslash f\}\} = \{\backslash n \backslash t \backslash r \backslash f\}$

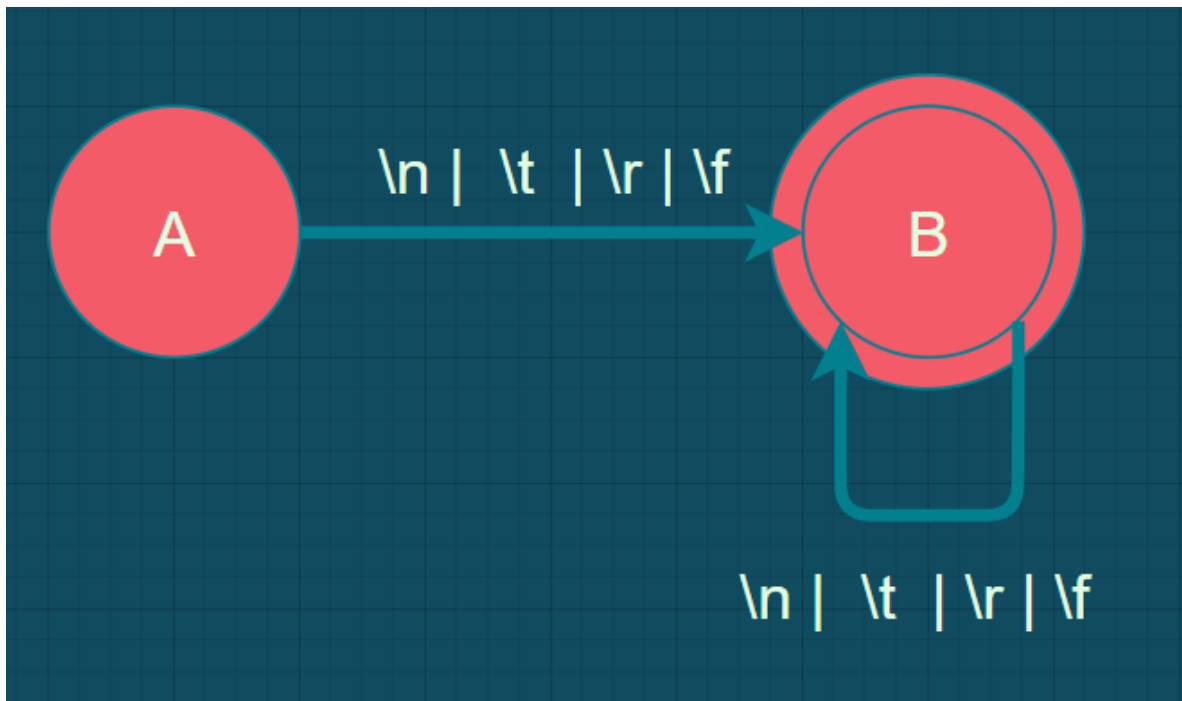
Expresión Regular: $(\backslash (n) (t) (r) (f)) ^ *$



FT	S	<code>\n \t \r \f</code>
s0	{s1}=A	(A, <code>\n \t \r \f</code>)=s2,s3,s4,s5,s6, s7 =B
s7	{s7}=B	(B, <code>\n \t \r \f</code>)= s7 =B

Reducido:

FT		\n \t \r \f
A		B
B		B

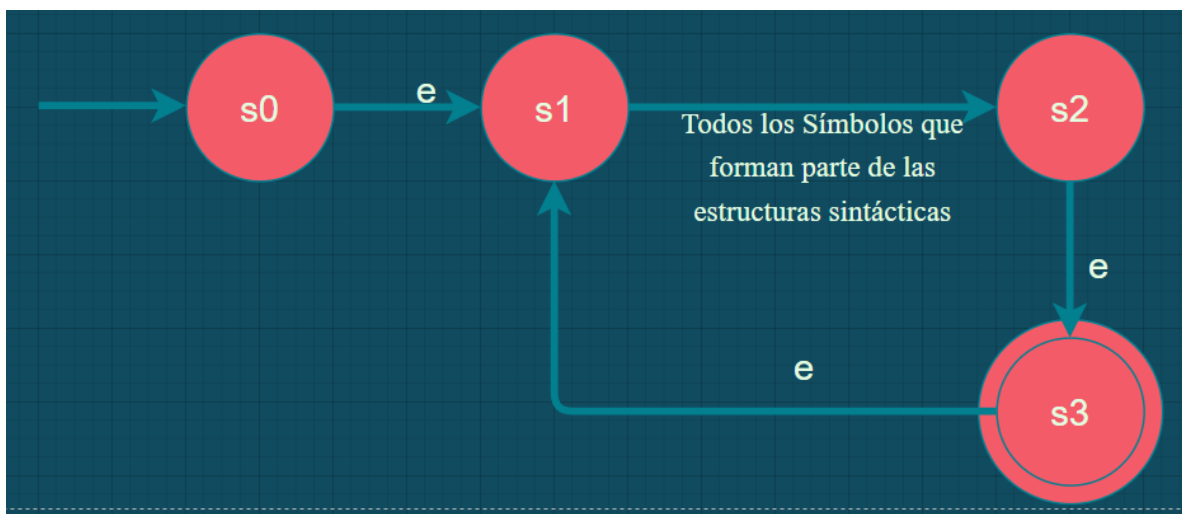


Símbolos De Estructuras Sintácticas.

Todos los Símbolos que forman parte de las estructuras sintácticas.

Alfabeto: {Todos los Símbolos que forman parte de las estructuras sintácticas} =
 {Todos los Símbolos que forman parte de las estructuras sintácticas}

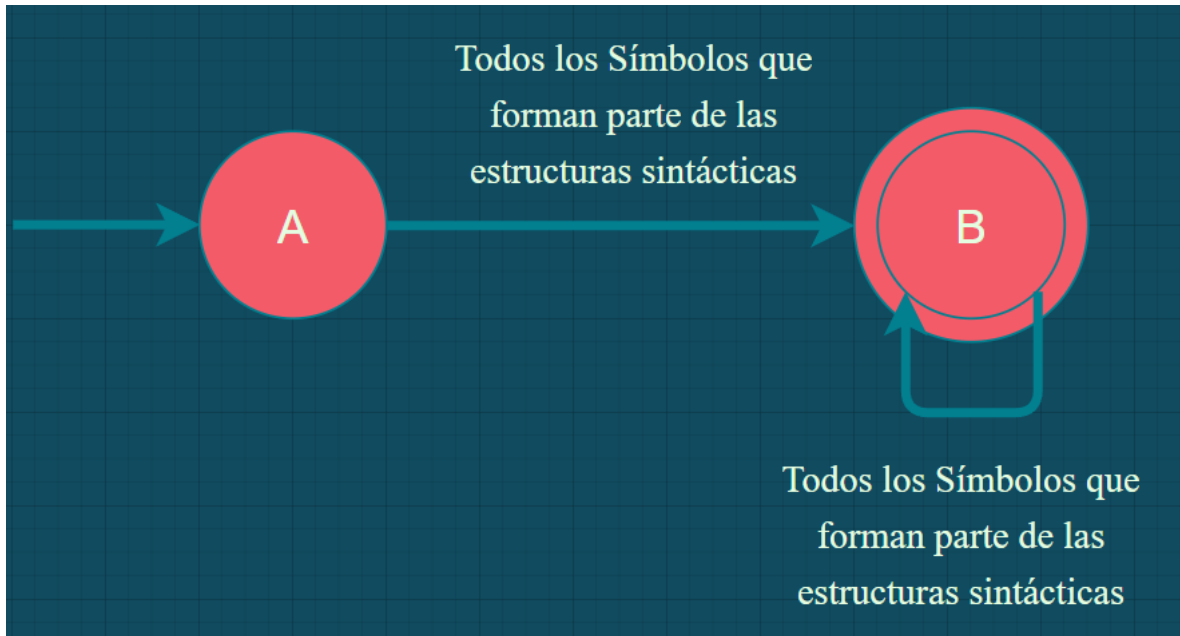
Expresión Regular: (Todos los Símbolos que forman parte de las estructuras sintácticas)
 *



FT	S	Sim. Estructura Sintáctica
s0	{s1}=A	(A,SES)=s2=B
s3,s1	{s2}=B	(B,SES)= s3 ,s1=B

Reducido:

FT	Sim. Estructura Sintactica
A	B
B	B



ANÁLISIS SINTÁCTICO.

Un Analizador Sintáctico o parser (viene del inglés: parse - analizar una cadena o texto en componentes sintácticos lógicos) es un programa que normalmente es parte de un compilador. El compilador se asegura de que el código se traduce correctamente a un lenguaje ejecutable. La tarea del analizador es, en este caso, la descomposición y transformación de las entradas en un formato utilizable para su posterior procesamiento. Se analiza una cadena de instrucciones en un lenguaje de programación y luego se descompone en sus componentes individuales.

Estructura ESCRITURA.

Está formada por la palabra reservada ESCRIBIR seguida de un literal, número o identificador, seguido del símbolo de fin de estructura el cual es la palabra reservada FIN.

Cuando el analizador sintáctico encuentre una estructura ESCRITURA correcta, debe escribir en una línea del archivo de salida lo siguiente:

- Palabra literal: lexema sin comillas.
- Para número: lexemas.
- Para identificador: el valor numérico almacenado.

Estructura Preliminar.

$E \Rightarrow \text{"ESCRIBIR"} S \text{"FIN"}$

$S \Rightarrow \text{Literal} \mid \text{Identificador} \mid \text{Número}$

Modificación.

$E \Rightarrow \text{"ESCRIBIR"} S \text{"FIN"}$

$S \Rightarrow \text{Literal} \mid N$

$N \Rightarrow \text{Número} \mid \text{Identificador}.$

Estructura REPETIR.

Está formada por la palabra reservada REPETIR seguida de un número entero positivo o de un identificador, seguido de la palabra reservada INICIAR seguido de cero o muchas estructuras ESCRITURA seguido del símbolo de fin de estructura el cual es la palabra reservada FIN.

Estructura Preliminar:

$R \Rightarrow \text{"REPETIR"} \ N \ \text{"INICIO"} \ G \ \text{"FIN"}$

$G \Rightarrow EG \mid \text{Epsilon}$

$F \Rightarrow E \mid \text{Epsilon}$

Estructura CONDICIONAL.

Está formada por la palabra reservada SI seguida de una de las palabras reservadas VERDADERO o FALSO, seguida de la palabra reservada ENTONCES, seguida de cero o una estructura ESCRITURA seguido del símbolo de fin de estructura el cual es la palabra reservada FIN. Si la estructura ESCRITURA está anidada en una estructura CONDICIONAL con palabra reservada VERDADERO entonces se debe escribir el contenido de la estructura ESCRITURA.

Estructura Preliminar:

$I \Rightarrow \text{"SI"} \ C \ \text{"ENTONCES"} \ F \ \text{"FIN"}$

$C \Rightarrow \text{"VERDADERO"} \mid \text{"FALSO"}$

$F \Rightarrow E \mid \text{Epsilon}$

$E \Rightarrow \text{"ESCRIBIR"} \ S \ \text{"FIN"}$

$S \Rightarrow \text{Literal} \mid N$

$N \Rightarrow \text{Identificador} \mid \text{Número}$

Estructura EXPRESIÓN.

Está formada por operaciones de suma y multiplicación entre números enteros e identificadores. Es posible la agrupación con paréntesis.

Estructura Preliminar:

$$X \Rightarrow V M D$$
$$M \Rightarrow * X \mid \text{Epsilon}$$
$$D \Rightarrow + X \mid \text{Epsilon}$$
$$V \Rightarrow \text{Identificador} \mid \text{Número} \mid \text{"("X")"}$$

Estructura ASIGNACIÓN.

Está formada por un identificador seguido de un signo igual (=) seguido de la estructura EXPRESIÓN seguida del símbolo de fin de estructura el cual es la palabra reservada FIN. El programa debe ser capaz de conocer el valor de un identificador en cualquier momento para poder ser usado en otras estructuras. Cuando se inicia el análisis sintáctico, el sistema debe preguntarle al usuario el nombre y ubicación del archivo de salida que contendrá los valores de las estructuras ESCRITURA.

IMPORTANTE: Se asume que en el texto de entrada los identificadores no se repiten. Se asume que en el texto de entrada los identificadores solo guardan valores de números enteros.

Estructura Preliminar:

$$A \Rightarrow \text{Identificador} \text{"="} X \text{"FIN"}$$
$$X \Rightarrow V M D$$
$$M \Rightarrow * X \mid \text{Epsilon}$$
$$D \Rightarrow + X \mid \text{Epsilon}$$
$$V \Rightarrow \text{Identificador} \mid \text{Número} \mid \text{"("X")"}$$

GLOBAL.

$L \Rightarrow E L \mid R L \mid I L \mid A L \mid \text{Epsilon}$

$E \Rightarrow \text{"ESCRIBIR"} S \text{"FIN"}$

$S \Rightarrow \text{Literal} \mid N$

$N \Rightarrow \text{Número} \mid \text{Identificador}$

$R \Rightarrow \text{"REPETIR"} N \text{"INICIO"} G \text{"FIN"}$

$G \Rightarrow E G \mid \text{Epsilon}$

$F \Rightarrow E \mid \text{Epsilon}$

$I \Rightarrow \text{"SI"} C \text{"ENCONTCE"} F \text{"FIN"}$

$C \Rightarrow \text{"VERDADERO"} \mid \text{"FALSO"}$

$A \Rightarrow \text{Identificador} \text{"="} X \text{"FIN"}$

$X \Rightarrow V M D$

$M \Rightarrow \text{"*"} X \mid \text{Epsilon}$

$D \Rightarrow \text{"+"} X \mid \text{Epsilon}$

$V \Rightarrow N \mid \text{"("X")"}$

PRIMEROS

No terminal	Primeros
L	{ "ESCRIBIR", "REPETIR", "SI", Identificador, Epsilon }
E	{ "ESCRIBIR" }
S	{ Literal, Número, Identificador }
N	{ Numero, Identificador }
R	{ "REPETIR" }
G	{ "ESCRIBIR", Epsilon }
F	{ "ESCRIBIR", Epsilon }
I	{ "SI" }
C	{ "VERDADERO", "FALSO" }
A	{ Identificador }
X	{ Número, Identificador, "(" }
M	{ "*", Epsilon }
D	{ "+", Epsilon }
V	{ Número, Identificador, "(" }

SIGUIENTES

No terminal	Siguientes
L	{ \$ }
E	{ \$, "ESCRIBIR", "REPETIR", "SI", Identificador }
S	{ "FIN" }
N	{ "FIN", "INICIO", "*", "+" }
R	{ \$, "ESCRIBIR", "REPETIR", "SI", Identificador }
G	{ "FIN" }
F	{ "FIN" }
I	{ \$, "ESCRIBIR", "REPETIR", "SI", Identificador }
C	{ "ENTONCES" }
A	{ \$, "ESCRIBIR", "REPETIR", "SI", Identificador }
X	{ "FIN", ")" }
M	{ "+", "FIN", ")" }
D	{ "FIN", ")" }
V	{ "*", "+", ")" , "FIN" }

TABLA DE ESTADOS.

.	"ESCRIBIR"	"REPETIR"	"SI"	Número	Literal	Identificador	"FIN"	"ε"	"ε"	"VERDADERO"	"FALSO"	"I"	"J"	\$
L	E L	R L	I L			A L								Epsilon
	"ESCRIBIR"													
S	S													
E	"FIN"													
S				N	Literal	N								
N				Número		Identificador								
		"REPETIR"												
		N												
		"INICIO"												
		G												
R		"FIN"												
G	E G									Epsilon				
F	E									Epsilon				
			"SI"											
			C											
			"ENTONCES"											
			F											
I			"FIN"											
C										"VERDADERO"	"FALSO"			
						Identificador								
						"ε"								
						X								
A						"FIN"								
X				V M D		V M D						V M D		
M							Epsilon	"ε" X	Epsilon				Epsilon	
D							Epsilon		"ε" X				Epsilon	
V				N		N						"(X)"		