

Especificación del Analizador Léxico

Aramis E. Matos, Lenier Gerena, Angel Berrios

2/28/2023

Índice I

- 1 Introduccion
- 2 Gramática de AVISMO
- 3 Especificación del Analizador Léxico
 - Tokens y Lexemas
 - Patrones
 - Atributos
- 4 Diseño del Analizador Léxico
 - Autómatas Finitos Deterministas
 - Estructura de Datos de la Tabla de Símbolos
- 5 Casos de Pruebas

Introduccion

Se estará definiendo los fundamentos léxicos del lenguaje AVISMO
(Narciso Farias, Rios, Hidrobo, & Vicuña, 2012)

En específico, los siguientes aspectos:

- Gramática del Lenguaje
(Reglas de producción para un lenguaje)
- Especificación del Analizador Léxico
- Diseño del Analizador Léxico

Gramática de AVISMO I

- $\langle \text{SENTENCIAS} \rangle ::= \langle \text{FIN_DE_LINEA} \rangle$
 $\langle \text{SENTENCIAS} \rangle \mid \langle \text{SENTENCIA} \rangle \langle \text{FIN_DE_LINEA} \rangle$
- $\langle \text{FIN_DE_LINEA} \rangle ::= ":" \mid ";"$
- $\langle \text{SENTENCIA} \rangle ::= \text{"defina"} \langle \text{ID} \rangle \text{"como"} \langle \text{TIPO} \rangle \mid$
 $\langle \text{ID} \rangle \text{"="} \langle \text{MODELO_MOLECULAR} \rangle \mid \langle \text{OPERACION} \rangle$
 $\text{"("} \langle \text{ID} \rangle \text{"}"$

Gramática de AVISMO II

- $\langle \text{ID} \rangle ::= \text{"A"} \mid \text{"B"} \mid \text{"C"} \mid \text{"D"} \mid \text{"E"} \mid \text{"F"} \mid \text{"G"} \mid \text{"H"} \mid \text{"I"} \mid \text{"J"} \mid \text{"K"} \mid \text{"L"} \mid \text{"M"} \mid \text{"N"} \mid \text{"O"} \mid \text{"P"} \mid \text{"Q"} \mid \text{"R"} \mid \text{"S"} \mid \text{"T"} \mid \text{"U"} \mid \text{"V"} \mid \text{"W"} \mid \text{"X"} \mid \text{"Y"} \mid \text{"Z"} \mid \text{"a"} \mid \text{"b"} \mid \text{"c"} \mid \text{"d"} \mid \text{"e"} \mid \text{"f"} \mid \text{"g"} \mid \text{"h"} \mid \text{"i"} \mid \text{"j"} \mid \text{"k"} \mid \text{"l"} \mid \text{"m"} \mid \text{"n"} \mid \text{"o"} \mid \text{"p"} \mid \text{"q"} \mid \text{"r"} \mid \text{"s"} \mid \text{"t"} \mid \text{"u"} \mid \text{"v"} \mid \text{"w"} \mid \text{"x"} \mid \text{"y"} \mid \text{"z"} \mid \langle \text{LETRA} \rangle \langle \text{IDCONT} \rangle$
- $\langle \text{IDCONT} \rangle ::= \text{"A"} \mid \text{"B"} \mid \text{"C"} \mid \text{"D"} \mid \text{"E"} \mid \text{"F"} \mid \text{"G"} \mid \text{"H"} \mid \text{"I"} \mid \text{"J"} \mid \text{"K"} \mid \text{"L"} \mid \text{"M"} \mid \text{"N"} \mid \text{"O"} \mid \text{"P"} \mid \text{"Q"} \mid \text{"R"} \mid \text{"S"} \mid \text{"T"} \mid \text{"U"} \mid \text{"V"} \mid \text{"W"} \mid \text{"X"} \mid \text{"Y"} \mid \text{"Z"} \mid \text{"a"} \mid \text{"b"} \mid \text{"c"} \mid \text{"d"} \mid \text{"e"} \mid \text{"f"} \mid \text{"g"} \mid \text{"h"} \mid \text{"i"} \mid \text{"j"} \mid \text{"k"} \mid \text{"l"} \mid \text{"m"} \mid \text{"n"} \mid \text{"o"} \mid \text{"p"} \mid \text{"q"} \mid \text{"r"} \mid \text{"s"} \mid \text{"t"} \mid \text{"u"} \mid \text{"v"} \mid \text{"w"} \mid \text{"x"} \mid \text{"y"} \mid \text{"z"} \mid \langle \text{LETRA} \rangle \langle \text{IDCONT} \rangle \mid$

Gramática de AVISMO III

"0" | "1" | "2" | "3" | "4" | "5" | "6" | "7" | "8" | "9" | <DIGITO>
<IDCONT>

■ <LETRA> ::= "A" | "B" | "C" | "D" | "E" | "F" | "G" | "H" | "I" |
"J" | "K" | "L" | "M" | "N" | "O" | "P" | "Q" | "R" | "S" | "T" | "U"
| "V" | "W" | "X" | "Y" | "Z" | "a" | "b" | "c" | "d" | "e" | "f" | "g" |
"h" | "i" | "j" | "k" | "l" | "m" | "n" | "o" | "p" | "q" | "r" | "s" | "t" |
"u" | "v" | "w" | "x" | "y" | "z"

Tokens y Lexemas

- Un lexema es una serie de caracteres que sigue un patrón
- Un *token* es un lexema que sigue un patrón y es asignado una categoría

<i>Token</i>	Lexema
<FIN_DE_LINEA>	;
<PALABRA_RESERVADA>	defina
<ID>	var1
<IDCONT>	01A
<ASIGNACION>	=
<LETRA>	b
<DIGITO>	6

Patrones

- Un patrón es una descripción de la forma que los lexemas de un token pueden tomar (Aho, Sethi, & Ullman, 1986)

<i>Token</i>	Patrones
<FIN_DE_LINEA>	; :
<PALABRA_RESERVADA>	defina como
<ID>	[A-Za-z] <LETRA> <IDCONT>
<IDCONT>	[A-Za-z] <LETRA> <IDCONT> [0-9] <DIGITO> <IDCONT>
<ASIGNACION>	=
<LETRA>	[A-Za-z]
<DIGITO>	[0-9]

Atributos

- Atributos son información adicional que describen las propiedades a un token

<i>Token</i>	Atributos
<FIN_DE_LINEA>	Indica fin de Línea
<PALABRA_RESERVADA>	Indica que el token no se puede utilizar como identificador
<ID>	Apuntador a la tabla de símbolos
<IDCONT>	Permite que los identificadores contengan números
<ASIGNACION>	Asigna un <MODELO_MOLECULAR> a un identificador
<LETRA>	Provee un terminal para <ID> y <IDCONT>
<DIGITO>	Provee un terminal para <ID>

Diseño del Analizador Léxico

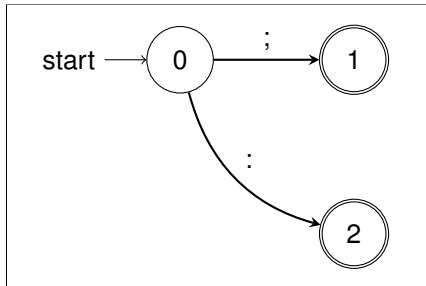


Figura: Automata del patrón
para el token
<FIN_DE_LINEA>

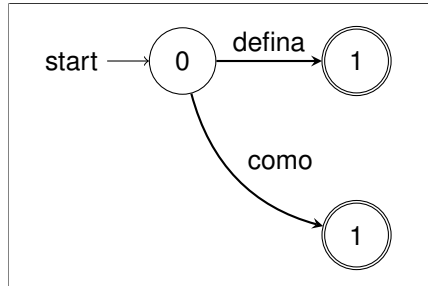


Figura: Automata del patrón
para el token
<PALABRAS_RESERVADA>

Diseño del Analizador Léxico

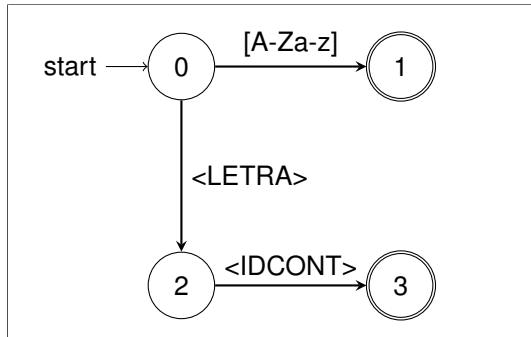


Figura: Automata del patrón para el token <ID>

Diseño del Analizador Léxico

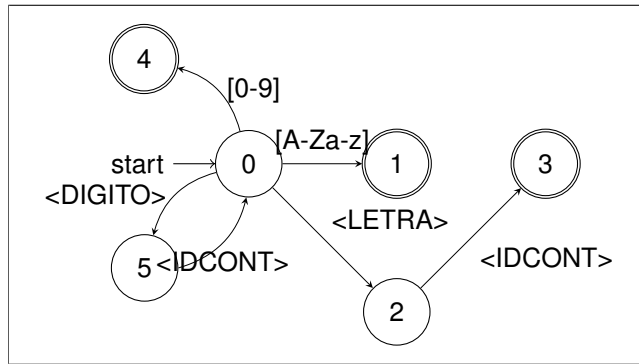


Figura: Automata del patrón para el token `<IDCONT>`

Diseño del Analizador Léxico

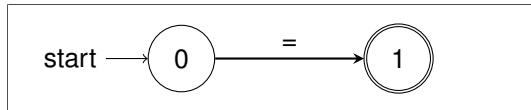


Figura: Automata del patrón para el token <ASIGNACION>

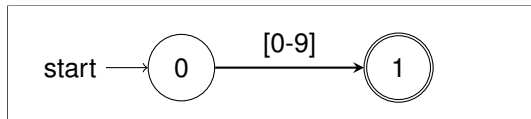


Figura: Automata del patrón para el token <LETRA>

Estructura de Datos de la Tabla de Símbolos

$\text{var1} \rightarrow \{\text{tipo} : \text{str}, \text{valor} : \text{"chungus"}, \text{esMutable} : \text{false}\}$

$\text{Big} \rightarrow \{\text{tipo} : \text{int}, \text{valor} : \text{"12"}, \text{esMutable} : \text{false}\}$

\vdots

Identificador \rightarrow **Atributos de Identificador**

Figura: Estructura de la Tabla de Símbolos, un Diccionario

Casos de Prueba

✗ Programas Léxicamente Incorrectos

- inicio → Error Léxico
defina a1 como modelo;
a1 = CH3CH(CH3)CH3;
fin

✓ Programas Léxicamente Correctos

- inicio → ID inicio
defina a1 como modelo;
a1 = CH3CH(CH3)CH3;
fin

Casos de Prueba

✗ Programas Léxicamente Incorrectos

- inicio defina a_1 como modelo; → Error Léxico
a1 = CH3CH(CH3)CH3;
fin

✓ Programas Léxicamente Correctos

- inicio defina a1 como modelo; → ID a1
a1 = CH3CH(CH3)CH3;
fin

Casos de Prueba

✗ Programas Léxicamente Incorrectos

- inicio defina a1 como
modelo → Error Léxico
a1 = CH3CH(CH3)CH3
→ Error Léxico
fin

✓ Programas Léxicamente Correctos

- inicio defina a1 como
modelo; → Fin de línea
a1 = CH3CH(CH3)CH3;
→ Fin de línea
fin

Referencias

Aho, A. V., Sethi, R., & Ullman, J. D. (1986). *Compilers, principles, techniques, and tools*. Reading, Mass: Addison-Wesley Pub. Co.

Narciso Farias, F., Rios, A., Hidrobo, F., & Vicuña, O. (2012, May). Una Gramática Libre de Contexto para el Lenguaje del Ambiente de Visualización Molecular - AVISMO..