	Carátula para entrega de prácticas	
Facultad de Ingeniería		Laboratorios de docencia

Laboratorio de Computación
Salas A y B

Profesor(a): René Adrián Dávila Pérez

Asignatura: Compiladores

Grupo: 4

Integrante(s): 315118894

317099681

315081143

313044270

Semestre: 2024-1

Fecha de entrega: 05/10/23

Observaciones:

CALIFICACIÓN: _____

Tabla de contenido

Introducción

Un analizador léxico o scanner, es un programa que analiza el texto fuente de un programa informático y lo divide en unidades más pequeñas, llamadas tokens. Estas unidades son los elementos básicos de un lenguaje de programación, como palabras reservadas, identificadores, operadores, literales, etc.

El analizador léxico es la primera etapa en el proceso de compilación de un programa informático. Su función es identificar los tokens que forman el código fuente y generar un flujo de tokens que será utilizado por el siguiente paso en el proceso de compilación, el analizador sintáctico.

Un scanner se puede desarrollar con distintos métodos:

- Con un autómata finito
- Con un programa a medida
- Utilizando una herramienta específica como Flex

En este trabajo se expondrá un analizador léxico (scanner) que procesa un lenguaje llamado “Tu_P_L”, un lenguaje en español. Para este proyecto se hará uso de la herramienta “PLY (Python Lex-Yacc)”, ya que hacer un autómata finito es una tarea increíblemente laboriosa y PLY proporciona las herramientas necesarias para construir de forma sencilla el analizador léxico.

Hipótesis

La implementación de un analizador léxico para el lenguaje específico propuesto permitirá analizar y dividir de manera efectiva y precisa el código fuente escrito en este lenguaje en una secuencia de tokens, que incluirán palabras reservadas, operadores aritméticos y relacionales, asignaciones, delimitadores, espacios, variables, cadenas, números y funciones, según las reglas de la gramática definida para este lenguaje.

Objetivos y Expectativas del Analizador Léxico:

1. **Reconocimiento de Palabras Reservadas:** El analizador léxico deberá identificar y clasificar correctamente las palabras reservadas, como "entero," "buleano," "flotante," "SI," "EntonCes," "Para," y "MiEntras." Estas palabras reservadas serán tokens fundamentales para definir la estructura y el flujo del programa.

2. **Identificación de Operadores Aritméticos y Relacionales:** El analizador léxico deberá detectar los operadores aritméticos ("+", "*", "-", "/") y los operadores relacionales ("<", ">", "=3", "E=)", "W!") cuando aparezcan en el código fuente.
3. **Manejo de Asignaciones:** El analizador léxico deberá ser capaz de reconocer y clasificar correctamente las asignaciones representadas por "<=".
4. **Delimitadores y Espacios:** Deberá identificar y gestionar adecuadamente los delimitadores (";" "(", ")") y los espacios en blanco ("\t," "\n," " ") que puedan aparecer en el código
5. **Identificación de Variables:** El analizador léxico deberá reconocer las variables representadas por el patrón ".|." y asignarles un token correspondiente.
6. **Reconocimiento de Cadenas:** Deberá identificar y clasificar las cadenas de caracteres, marcadas por "CaD," como tokens válidos.
7. **Detección de Números:** El analizador léxico deberá identificar secuencias de dígitos ([0-9]+) y considerarlas como tokens de números válidos.

Se espera que el analizador léxico funcione de manera eficiente incluso para programas largos y complejos, procesando el código fuente de manera rápida y sin consumo excesivo de recursos. El analizador léxico propuesto tiene como objetivo fundamental dividir el código fuente escrito en el lenguaje específico en una secuencia de tokens válidos. Su éxito se medirá por su capacidad para reconocer y clasificar los elementos léxicos del lenguaje según las reglas definidas

Desarrollo

Para la construcción del lenguaje "Tu_P_L" primero se tiene que definir cómo será el lenguaje, este lenguaje se escogió por parte de los integrantes del equipo y se propuso lo siguiente

Lenguaje

Primero se definió los tokens o elementos léxicos del lenguaje. Esto incluye palabras reservadas, operadores, delimitadores, tipos de datos, identificadores, números y cualquier otro elemento que sea significativo en el lenguaje. Cada token se define utilizando expresiones regulares o patrones que describen cómo se ven y cómo se deben reconocer.

Palabras reservadas:

- "entero"
- "buleano"
- "flotante"
- "SI"
- "EntonCes"
- "Para"
- "MiEntras"

Estas son palabras propias del lenguaje que serán utilizadas para reconocer sentencias y ciertas estructuras del lenguaje y no con otros fines, por ejemplo como identificadores.

Operadores aritméticos:

- "+"
- "*"
- "-"
- "/"

Tu_P_Lenguaje permitirá realizar las operaciones aritméticas básicas. Estos son los tokens que se utilizarán para poder formar dichas expresiones.

Operadores relacionales:

- "<" (menor que)
- ">" (mayor que)
- "(=3" (menor o igual que)
- "E=)" (mayor o igual que)
- "W!" (diferente de)

Estos tokens definen las distintas comparaciones que se pueden hacer entre los operandos.

Asignación

- "<=" (igual a)

Este token es un operador que se utilizará para asignar valores a las variables.

Delimitadores:

- ";"
- "("
- ")"

Sirven para agrupar elementos o delimitar el fin de alguna estructura.

Espacios

- "\t"
- "\n"
- " "

Para los espacios en blanco (tabuladores y espacios) y los saltos de línea. En Tu_P_Lenguaje estos caracteres no tienen significado especial para el lenguaje y deberá ignorarlos.

Variable

- ".|."[a-ZA-Z0-9]

Sirve para identificar las cadenas válidas que pueden ser utilizadas como identificadores de variables.

Cadena

- "CaD"[A-Za-z0-9]";"

Se planea que el lenguaje tenga soporte de cadenas. Este patrón sirve para identificar las cadenas válidas.

Número

- "[0-9]+"

Este para reconocer número, especificamente, numeros enteros. Esta versión del lenguaje no tendrá soporte para números con valores decimales.

Función

- ":-*"

Token para reconocer funciones válidas del lenguaje. Sería como el equivalente a la palabra "def" que se utiliza en Python.

Uso de Herramientas

PLY es un acrónimo de "Python Lex-Yacc", que es una biblioteca de Python utilizada para el análisis léxico y sintáctico. PLY proporciona una forma de crear analizadores léxicos y sintácticos. Un analizador léxico (también conocido como escáner) toma una entrada (como el código fuente de un programa) y la divide en tokens, mientras que un analizador sintáctico (o parser) toma estos tokens y los utiliza para construir una estructura de datos, como un árbol de sintaxis abstracta.

El módulo lex.py que se usará en este proyecto se utiliza para dividir el texto de entrada en una colección de tokens especificados por un conjunto de reglas de expresiones regulares.

El formato de un archivo de entrada para PLY tiene la siguiente estructura

```
Definiciones
%%
Reglas
%%
Código del usuario
```

En la seccion de definiciones se colocan, entre otras cosas, declaraciones de nombres sencillos que tienen la finalidad de simplificar la especificación del analizador léxico.[2]

En la sección de reglas es donde se escriben los patrones (expresiones regulares) que va a reconocer asi como las acciones que se van a realizar una vez que el analizador encuentre alguna cadena que coincida con dichos patrones.

Las expresiones regulares de nuestros tokens simples son las siguientes:

```
t_SUMA = r'\+'

t_MULTIPLICACION = r'\*'

t_MENOS = r'\-'

t_ENTRE = r'/'
```

```
t_MENOR_QUE = r'<'
t_MAYOR_QUE = r'>'
t_MENOR_O_IGUAL_QUE = r'<=3'

t_MAYOR_O_IGUAL_QUE = r'=>\*'

t_DIFERENTE_DE = r'{\|}'

t_ASIGNACION = r'<='

t_IGUAL_A = r'='

t_DELIMITADOR = r";"

t_PARENTESIS_IZQ = r'\('

t_PARENTESIS_DER = r'\)'

t_ignore = ' \t
```

Expresión regular para comentarios entre comillas:

```
t_COMENTARIO = r'"[^"]*"'
```

Para ignorar caracteres en blanco:

```
t_ignore = ' \t'
```

Definimos las Reglas

Salto de linea:

```
def t_SaltoLinea(t):
    r'\n+'
    t.lexer.lineno += len(t.value)
```

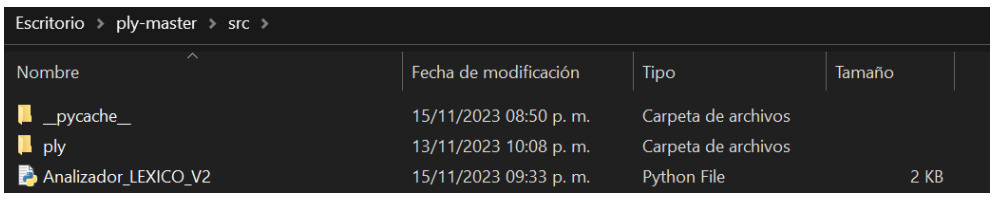
Manejo de errores:

```
def t_error(t):
    print("Carácter ilegal: ", t.value[0])
    t.lexer.skip(1)
```

Ejecución

Para la ejecución del código se necesita de PLY, el cual se puede descargar de https://www.dabeaz.com/ply/ply.html

Se deberá ubicar nuestro código en la carpeta “src”.



Una vez ubicado en esta carpeta se puede hacer la ejecución. A continuación se hará una ejecución con la siguiente entrada:

```
data = '''
+
*
-
/
<
>
(
)
<=3
=>*
{|}
.|.VARIABLE1    <=  4656 ;
entero ;
SI .|.x <= 10 ;
'''
```

La salida de esta ejecución es la siguiente:

```
= RESTART: C:\Users\spart\OneDrive\Escritorio\ply-master\src\Analizador_LEXICO_V2.py
LexToken(SUMA,'+',2,1)
Tipo: SUMA, Valor: +
LexToken(MULTIPLICACION,'*',3,3)
Tipo: MULTIPLICACION, Valor: *
LexToken(MENOS,'-',4,5)
Tipo: MENOS, Valor: -
LexToken(ENTRE,'/',5,7)
Tipo: ENTRE, Valor: /
LexToken(MENOR_QUE,'<',6,9)
Tipo: MENOR_QUE, Valor: <
LexToken(MAYOR_QUE,'>',7,11)
Tipo: MAYOR_QUE, Valor: >
LexToken(PARENTESIS_IZQ,'(',8,13)
Tipo: PARENTESIS_IZQ, Valor: (
LexToken(MENOR_O_IGUAL_QUE,'<=3',9,15)
Tipo: MENOR_O_IGUAL_QUE, Valor: <=3
LexToken(DIFERENTE_DE,'{|}',10,19)
Tipo: DIFERENTE_DE, Valor: {|}
LexToken(ID,'.|.VARIABLE1',11,23)
Tipo: ID, Valor: .|.VARIABLE1
LexToken(ASIGNACION,'<=',11,39)
Tipo: ASIGNACION, Valor: <=
LexToken(NUMERO,'4656',11,43)
Tipo: NUMERO, Valor: 4656
LexToken(DELIMITADOR,';',11,48)
Tipo: DELIMITADOR, Valor: ;
LexToken(ENTERO,'entero',12,50)
Tipo: ENTERO, Valor: entero
LexToken(DELIMITADOR,';',12,57)
Tipo: DELIMITADOR, Valor: ;
LexToken(SI,'SI',13,59)
Tipo: SI, Valor: SI
LexToken(ID,'.|.x',13,62)
Tipo: ID, Valor: .|.x
LexToken(ASIGNACION,'<=',13,67)
Tipo: ASIGNACION, Valor: <=
LexToken(NUMERO,'10',13,70)
Tipo: NUMERO, Valor: 10
LexToken(DELIMITADOR,';',13,73)
Tipo: DELIMITADOR, Valor: ;
>>>
```

Conclusiones

El desarrollo de un analizador léxico utilizando la herramienta PLY es un paso fundamental en la construcción de un compilador, intérprete o procesador de lenguajes de programación. El analizador léxico se encarga de escanear y dividir el código fuente en una secuencia de tokens, lo que sienta las bases para una interpretación o compilación precisa.

Algunos de los aspectos clave a tener en cuenta en el desarrollo de un analizador léxico con PLY incluyen la definición de reglas para reconocer patrones léxicos, la asignación de valores, la devolución de tokens, y la construcción de una gramática, así como la gestión de errores. Además, es importante que el analizador sea eficiente y capaz de manejar de manera adecuada el flujo de entrada.

En conclusión un analizador léxico bien desarrollado y probado es crucial para el éxito de proyectos de compilación o interpretación, ya que garantiza que el código fuente se interprete o compile de manera correcta y eficiente. Además, una documentación adecuada del analizador léxico facilita su uso y mantenimiento en proyectos futuros.

Referencias

[1] D. Beazley, “PLY (Python Lex-Yacc),” Ply (python lex-yacc), https://www.dabeaz.com/ply/ (accessed Oct. 13, 2023).

[2] Paxson Vern. "Flex, version 2.5". USA: THe Regents of the Univerity of California, 1990.