Gerardo Guillermo Garza Tamez

A01380899  06/06/22

Compilador Veigar

Índice

[**Descripción del proyecto** 2](#_Toc88688341)

[**Propósito y Alcance del proyecto** 2](#_Toc88688342)

[**Análisis de requerimientos y descripción de los principios test Cases** 2](#_Toc88688343)

[**Descripción del proceso** 2](#_Toc88688344)

[Descripción del lenguaje 3](#_Toc88688345)

[Descripción del compilador 3](#_Toc88688346)

[Descripción del análisis de léxico 3](#_Toc88688347)

[**Descripción del análisis de sintaxis** 5](#_Toc88688348)

[**Descripción de generación de código intermedio y análisis Semántico** 9](#_Toc88688349)

[**Descripción de la Máquina virtual** 9](#_Toc88688350)

[**Pruebas del funcionamiento del lenguaje** 10](#_Toc88688351)

[**Documentación del código del proyecto** 10](#_Toc88688352)

# **Descripción del proyecto**

## **Propósito y Alcance del proyecto**

El propósito de esto proyecto fue el realizar un compilador comprendiendo todas las partes que este conforma, desde como se analizan los elementos estáticos que serian los tokens y como se debe estructurar generando un código intermedio para después formar una estructura para que pueda ser analizado por una maquina virtual y poder hacer las diferentes operaciones.

## **Análisis de requerimientos y descripción de los principios test Cases**

Como requerimientos funcionales serían los siguientes:

* El compilador debía poder reconocer diferentes tokens tales como if, else, program, int , float, print, return y otros mas especiales como circulo, punto, línea, entre otros
* El programa debe ser capaz de hacer las operaciones básicas como sumar, restar, multiplicar y dividir
* Debe guardar las funciones en algún arreglo
* Debe poder guardar las variables
* Debe generar cuadruplos

Como requerimientos no funcionales serian:

* Se debe guardar en la memoria de manera efectiva

## **Descripción del proceso**

Para el desarrollo de este proceso se inició generando las gramáticas e identificando cuales serian los caracteres clave, así como las palabras reservadas para generar los tokens.

En las siguientes semanas se comenzó a trabajar ya en el código se creó el léxico en Python todo el proyecto se desarrolló en Python y se creo la base del parser e interprete.

Mientras iba pasando las semanas se iban incluyendo las clases extra que se fueran a necesitar como la tabla de variables, cubo semántico etc.

Se encontraron algunas cosas que estaban incorrectas debido a que la investigación antes realizada no era exactamente lo que se necesitaba y en la adaptación ocasiono fallos en la lógica.

Como reflexión seria darle su tiempo y respeto a lo que se necesita si ocupas ayuda con algo no te quedes callado muévete y búscala con compañeros o profesores.

A continuación, está la liga de GitHub: https://github.com/Gera-Garza/compilador/tree/Final

# **Descripción del lenguaje**

Nombre del lenguaje: Veigar

Este lenguaje genera los cuádruplos básicos de suma resta multiplicación etc. A su vez puede realizar operaciones como condicionales, también tiene dos tipos de iteraciones el for loop y el while loop, el lenguaje que interpreta es muy similar a C con pocas diferencias como que solo se puede declarar variables globales al principio después de indicar el nombre del programa o declarar las variables locales al principio de cada función ya sea el main o cualquier otra función, permitirá arreglos pero de máximo 1 dimencion, no hay variables tipo string ni double, long o bool pero si es capaz de interpretar FALSE, TRUE y condicionales.

# Descripción del compilador

Se género en Windows con un procesador dual core y 16Gb de ram, el lenguaje en el que se desarrollo el compilador fue Python y de utilerías se usaron las librerías de ply para el léxico y parser, la librería de os para poder navegar en el sistema, la librería de códecs para abrir un archivo y la librería de re para generar las expresiones regulares

## Descripción del análisis de léxico

se guardaron las siguientes palabras como tokens:

reservadas = {

    'PROGRAM': 'program',

    'ID': 'id',

    'INT': 'int',

    'FLOAT': 'float',

    'CHAR': 'char',

    'IF': 'if',

    'ELSE': 'else',

    'READ': 'read',

    'VOID': 'void',

    'RETURN': 'return',

    'FOR': 'for',

    'WHILE': 'while',

    'AVG': 'avg',

    'MEDIAN': 'median',

    'MODE': 'mode',

    'MAX': 'max',

    'MIN': 'min',

    'STDEV': 'stdev',

    'HIST': 'hist',

    'COLOR': 'color',

    'BINS': 'bins',

    'EDGECOLOR': 'edgecolor',

    'PLT': 'plt',

    'XLABEL': 'xlabel',

    'YLABEL': 'ylabel',

    'TITLE': 'title',

    'SHOW': 'show',

    'BEGIN': 'begin',

    'ENDFUN': 'endfun'

}

tokens = reservadas + [

# ; { } , = ( ) [ ]

'SEMICOLON','L\_BRACE', 'R\_BRACE',

'COMMA', 'ASSIGN', 'L\_PARENTHESIS','R\_PARENTHESIS',

'L\_BRACKET','R\_BRACKET',

#operators + - \* /

'PLUS', 'MINUS', 'MULT', 'DIV',

#bool > < != == && ||

'GREATER\_THAN', 'LESS\_THAN', 'DIFF\_THAN',

'EQUALS\_TO', 'AND', 'OR',

#Constants

'CONST\_ID', 'CONST\_INT', 'CONST\_FLOAT', ‘CONST\_CHAR’,’LETRERO’

]

Para estos se generaron las siguientes expresiones regulares:

t\_ignore = '\t\n'

t\_ignore\_space = '\s'

t\_SEMICOLON = r';'

t\_L\_BRACE = r'\['

t\_R\_BRACE = r'\]'

t\_COMMA = r','

t\_ASSIGN = r'='

t\_L\_PARENTHESIS = r'\('

t\_R\_PARENTHESIS = r'\)'

t\_L\_BRACKET = r'\{'

t\_R\_BRACKET = r'\}'

t\_PLUS = r'\+'

t\_MINUS = r'\-'

t\_MULT = r'\\*'

t\_DIV = r'/'

t\_GRATHER\_THAN = r'\>'

t\_LESS\_THAN = r'\<'

t\_DIFF\_THAN = r'\!='

t\_EQUALS\_TO = r'=='

t\_AND = r'&&'

t\_OR = r'\|\|'

t\_CONST\_ID = r'[a-zA-Z\_][a-zA-Z\_0-9]\*'

t\_CONST\_INT =  r'\d+'

t\_CONST\_FLOAT= r'\d\.\d+'

t\_CONST\_CHAR = r'[a-zA-Z\_0-9]'

t\_LETREROS=  r'[a-zA-Z\_][a-zA-Z\_0-9]\*'

## **Descripción del análisis de sintaxis**

Para la generación de la sintaxis se tuvo que evaluar la estructura de como es que serian interpretado en el lenguaje y se generaron los siguientes diagramas

Diagrama

Descripción generada automáticamente

Diagrama

Descripción generada automáticamente

Diagrama, Esquemático

Descripción generada automáticamenteDiagrama

Descripción generada automáticamente

Diagrama

Descripción generada automáticamente

Diagrama

Descripción generada automáticamente

Diagrama

Descripción generada automáticamente

Diagrama

Descripción generada automáticamente

Diagrama

Descripción generada automáticamente

Diagrama

Descripción generada automáticamente

Diagrama

Descripción generada automáticamente

Diagrama

Descripción generada automáticamente

Imagen en blanco y negro

Descripción generada automáticamente con confianza media

Diagrama

Descripción generada automáticamente

Diagrama

Descripción generada automáticamente

Diagrama

Descripción generada automáticamente

Diagrama

Descripción generada automáticamente

Imagen en blanco y negro

Descripción generada automáticamente con confianza media

Imagen en blanco y negro

Descripción generada automáticamente con confianza baja

Diagrama

Descripción generada automáticamente

## **Descripción de generación de código intermedio y análisis Semántico**

Para la generación de código intermedio se generaron cuadruplos y se asignaron ciertas direcciones para la memoria siendo los caracteres y tokens reservados son números fijos que van desde el [- al - ] para palabras reservadas como int, float char, read, for, if etc. Y para los caracteres especiales como “,”, “;”, “=”,”\*”, “+” etc se les asigno un numero del [- al -] siendo estos caracteres que nunca podrán cambiar de posición estos fueron los números que les asigne, separe la memoria en 10 secciones la primera seria para los enteros que designe los números 500 en adelante para los enteros los enteros temporales serían los 600 en adelante los números flotantes del 750 en adelante cabe recalcar que para los números enteros flotantes separe mas ya que lo use como los temporales los números flotantes temporales son a partir del 850, los caracteres en el 900 los caracteres temporales en el 1000 y los strings o letreros en 1100 y los temporales de estos en el 1200.

Use la lista que seria lo mismo que usar una pila ya que era algo sencillo y al tenerlos todos separados por cierta cantidad los vuelve fácil de buscar o filtrar.

## **Descripción de la Máquina virtual**

La maquina virtual es el programa que se encargara de analizar los cuadruplos generados por el parser así como tener acceso a la tabla de funciones y la tabla de variables para obtener los datos de cada tabla para poder realizar las operaciones usa la misma “clase” memoria como simulación de la memoria requerida del interprete a la computadora.

## **Pruebas del funcionamiento del lenguaje**

## **Documentación del código del proyecto**

def t\_vars(t):

    global numInt, numFloat, numTxt

*if* t[1] == 'INT':

        tabla.addVar(t[1], t[2], numInt)

        memo.enteros.append(numInt)

        numInt += 1

*elif* t[1] == 'FLOAT':

        tabla.addVar(t[1], t[2], numFloat)

        memo.flotantes.append(numFloat)

        numFloat += 1

*else*:

        tabla.addVar(t[1], t[2], numTxt)

        memo.text.append(numTxt)

        numTxt += 1

def t\_varst(t):

    global numIntT, numFloatT, numtxtT

*if* t[1] == 'INT':

        tabla.addVar(t[1], t[2], numIntT)

        memo.ent\_temp.append(numIntT)

        numIntT += 1

*elif* t[1] == 'FLOAT':

        tabla.addVar(t[1], t[2], numFloatT)

        memo.flot\_temp.append(numFloatT)

        numFloatT += 1

*else*:

        tabla.addVar(t[1], t[2], numtxtT)

        memo.text\_temp.append(numtxtT)

        numtxtT += 1