

Diseño de compiladores

Maestra Elda Quiroga

Proyecto: Shinto

Sergio González Sifuentes

A00821229



06/06/2022

INDICE

Descripción del proyecto…………………………..1

Descripción del lenguaje…………………………..11

Descripción del compilador………………………..12

Descripción de la maquina virtual…………………18

Pruebas de funcionamiento………………………..18

**Descripción del Proyecto:**

El propósito de este lenguaje es evaluar los conocimientos del alumno en base a lo aprendido en la clase, el objetivo fue desarrollar un lenguaje básico con estatutos, variables, funciones, lectura, escritura y expresiones. Usando todo lo aprendido en clase: léxico, semántica, sintáxis, creación de código y ejecución

|  |  |
| --- | --- |
| **Caso de uso:**  UU001 | **Actor primario:**  Usuario |
| **Descripción:**  Se asignará un dato a una variable | |
| **Condición de éxito**  La variable es asignada con éxito | |
| **Workflow:**   1. Se ejecuta el compilador 2. Se definirá dentro de una función o globalmente una variable con tipo y nombre de la sig manera: var (tipo de dato) (nombre de la variable) ; (Pueden declarararse n variables del mismo tipo con diferentes nombres en una misma linea 3. El compilador crea una variable con esos datos, creandole una dirección y almacenándola en el directorio de variables como una instancia de la clase Variable | |
| **Precondición**  Ninguna | |
| **Restricciones y especificación**  No debe existir una variable con ese nombre | |

|  |  |
| --- | --- |
| **Caso de uso:**  UU002 | **Actor primario:**  Usuario |
| **Descripción:**  Se realizarán operaciones aritmeticas | |
| **Condición de éxito**  La respuesta mostrara un resultado siguiendo la jerarquía de operaciones | |
| **Workflow:**   1. Se ejecuta el compilador 2. Se escribe una expresión con sintáxis correcta 3. El compilador va guardando las operaciones en el stack y los operadores en otro para poder formar los cuadruplos correspondientes | |
| **Precondición**  Ninguna | |
| **Restricciones y especificación**  Ninguna | |

|  |  |
| --- | --- |
| **Caso de uso:**  UU003 | **Actor primario:**  Usuario |
| **Descripción:**  Se declarará una función | |
| **Condición de éxito**  El compilador ejecuta con éxito la función | |
| **Workflow:**   1. Se crea una función de la siguiente manera: (tipo de dato) function (nombre de la función) ((parámetros si es que los hay escritos como (nombre del parámetro, tipo de dato)) ) { (estatutos si es que los hay) } 2. Se crea la función y se agrega al directorio de funciones con una dirección 3. Se agregan los parámetros al directorio de funciones y se le agregan al diccionario de parámetros dentro de su función correspondiente 4. Se agregan las variables que se declaran en el inicio de la función al directorio de variables y se le agregan al diccionario de parámetros 5. Se crean los cuadruplos de los statements dentro de la función 6. Cerramos la función y si no hubo ningún problema, seguimos | |
| **Precondición**  Ninguna | |
| **Restricciones y especificación**  Ninguna | |

|  |  |
| --- | --- |
| **Caso de uso:**  UU004 | **Actor primario:**  Usuario |
| **Descripción:**  Se regresara un valor en una función | |
| **Condición de éxito**  Se regresa con éxito el valor esperado | |
| **Workflow:**   1. Se ejecuta el compilador 2. Se teclea “fun multiply(x, y); return x \* y end” 3. Se teclea “multiply(3, 5)” 4. El programa regresa “15” | |
| **Precondición**  Ninguna | |
| **Restricciones y especificación**  Ninguna | |

|  |  |
| --- | --- |
| **Caso de uso:**  UU005 | **Actor primario:**  Usuario |
| **Descripción:**  Se leerá un número y se guardará en una variable | |
| **Condición de éxito**  Al teclear la variable se regresará lo que se había tecleado anteriormente | |
| **Workflow:**   1. Se ejecuta el compilador 2. Se teclea “var str = input()” 3. El programa pedirá una string 4. Se teclea “Hello World!” 5. Se teclea “str” 6. El programa responderá con “Hello World!” | |
| **Precondición**  Ninguna | |
| **Restricciones y especificación**  Ninguna | |

Primero se desarrollo junto con sly la generación de tokens, la detección de palabras, números, etc. Además de retornar los valores de los int, float y booleans para después poder trabajar con ellos más adelante.

Se definieron los siguientes tokens:

Text

Description automatically generated

Además, se definió un diccionario para reservar palabras del lenguaje, palabras clave que se muestran a continuación:

Text

Description automatically generated

También, sly ayudó a poder definir literales y que no se tuviesen que generar tokens para ellos, los cuales se muestran a continuación:

A screenshot of a computer

Description automatically generated with medium confidence

También se definieron las reglas para generar los distintos tokens y que se identificaran correctamente:

Text

Description automatically generated

Primero se hizo una clase de Tipos de Dato que me ayudaría a trabajar fácilmente con la manipulación de datos, después con eso pude empezar a trabajar con el Cubo Semántico en el que podríamos hacer match con cada tipo de dato y el match regresaría un tipo de dato y si no fuera posible, se regresaría un tipo de dato inválido.

Antes de comenzar a trabajar de lleno en la gramática decidí empezar a trabajar en como se almacenarían los datos que se van tomando y generando.

Se hicieron 3 clases de tablas global, local y de constantes, como sus nombres lo dicen global sería para los datos que fueron inicializados en el scope global, local serían los que fueron definidos dentro de una función y la tabla de constantes sería para las constantes que vayamos usando y que necesitemos guardar.

Antes de poder almacenar en ellas, se tenían que delimitar la cantidad de almacenamiento que tendría cada diccionario de cada tabla así que se hizo una clase Delimitation que almacenaría el lugar en el que empieza cada tipo de dato de cada tabla y el numero de entradas de ese tipo que vamos llenando.

Para cada tabla se hicieron 4 diccionarios para int, float, string y booleans en el que cada uno haría la verificación de que no sea repetido antes de insertarlo en el diccionario.

Después se hizo una clase llamada Variables con los siguientes atributos:

Text

Description automatically generated

Y también una clase llamada Función con los siguientes atributos:

Text

Description automatically generated

La cual también incluye un diccionario de parámetros y un diccionario de variables

Después se hizo una clase para la Jerarquía de operadores para poder hacer las operaciones de manera ordenada y que resulten bien a la hora de meterlas al stack de operadores y para poder identificarlas más fácilmente.

Ahora, se empezó a trabajar en la clase de Memoria de Ejecución donde se escribiría donde va cada variable dependiendo de su scope (global, constante o local) y su tipo de dato (int, float, string y bool), además un método para convertir el dato almacenado en su tipo de dato correspondiente, se consigue los números de espacios que se necesitan en cada tipo de scope que vamos a conseguir sacando la longitud del diccionario que crearemos dentro del Parser que aun no codificamos pero para irnos haciendo la idea de como trabajaremos, un método para sacar un valor de una dirección en particular y por último un método para guardar un valor en una dirección.

Después, trabajaremos en el VM, donde tendremos una instancia de la memoria de ejecución, una instancia de la jerarquía de operadores un arreglo vacio de los cuadruplos y una variable ip (instruction pointer) para poder movernos atraves en la ejecución, al iniciar se recibirá un diccionario que el parser generará donde recibiremos un diccionario de globales, constantes, locales y los cuadruplos, después se ejecutara cada uno de ellos. En el método de ejecución y resolución de cuadruplos se ira identificando el tipo de operación con ayuda de la clase de Jerarquía para identificarlos, reconocer la operación, realizarla, guardar el dato correspondiente en la dirección solicitada y por último aumentar la dirección del ip para seguir o en caso de tener un goto o un gotof ir al num de cuádruplo solicitado.

Ahora para el parser que vendría siendo lo más importante a mi parecer se inicializa con las siguientes variables:

Text

Description automatically generated

Un debugfile que pertenece al sly y le damos el nombre de un archivo para que lo cree o lo reescriba si es que ya existe donde pondrá el camino que se utilizó en la gramática así como las reglas, meramente para hacer debugging, un arreglo de tokens que son los que definimos en el leer, un ERROR\_FLAG para identificar errores, un stack de dimensiones para manejar los arrays, un stack de parámetros para cuando trabajemos con las funciones y los vayamos a adjuntar a las funciones, un stack de variables para declarar variables y un stack de gvars para declarar variables globales, un arreglo de call\_params para poder trabajar con llamadas de funciones, un contador de parámetros igualmente para comparar el num de parámetros llamados con los declarados y un contador en general.

Text

Description automatically generated

También tenemos todas estas instancias de lo que se trabajó previamente para poder trabajar con ellos.

Para cada gramática se definirán los tokens que se deben utilizar y funciones que necesitan vacíos para poder hacer cálculos y generar los cuadruplos.

Al final se parseara la data para hacer un diccionario que pueda ser utilizado por la VM.

A screenshot of a computer

Description automatically generated with medium confidence

Errores:

Los errores que se pueden encontrar son:

Al hacer un cuádruplo se tenga un error de que el polish vector esté vacío

**Descripción del Compilador:**

Se utilizó la pc personal para trabajar en este compilador, usando Visual Studio Code, con el lenguaje Python y varias extensiones de VSCode como Python, pylint, github. Se utilizó Sly que es una implementación de las herramientas lex y yacc con documentación decente y métodos muy útiles para desarrollar el Parser y el Lexer del lenguaje usando un algoritmo LALR(1)

Análisis del Léxico:

A picture containing table

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

Análisis de Sintáxis:

Grammar:

Rule 0 S' -> program

Rule 1 program -> PROG ID check\_program ; gvars store\_gvars functions gvars store\_gvars main

Rule 2 gvars -> VAR datatype gvarids store\_gtype ; gvars

Rule 3 gvars -> <empty>

Rule 4 gvarids -> ID

Rule 5 gvarids -> ID , gvarids

Rule 6 store\_gvars -> <empty>

Rule 7 store\_gtype -> <empty>

Rule 8 vars -> VAR datatype varids store\_type ;

Rule 9 vars -> <empty>

Rule 10 varids -> ID

Rule 11 varids -> ID , varids

Rule 12 store\_type -> <empty>

Rule 13 functions -> datatype FUNC ID ( params ) store\_funcv store\_params store\_init\_quad { vars store\_local\_vars funcontent } close\_func functions

Rule 14 functions -> <empty>

Rule 15 funcontent -> <empty>

Rule 16 funcontent -> statement funcontent

Rule 17 params -> ID : datatype , params

Rule 18 params -> ID : datatype

Rule 19 params -> <empty>

Rule 20 main -> FUNC MAIN ( ) store\_funcm { vars store\_mainv maincontent }

Rule 21 maincontent -> <empty>

Rule 22 maincontent -> statement maincontent

Rule 23 statement -> loop unload\_pv

Rule 24 statement -> input unload\_pv

Rule 25 statement -> output unload\_pv

Rule 26 statement -> ifelse unload\_pv

Rule 27 statement -> returns unload\_pv ;

Rule 28 statement -> expr ;

Rule 29 statement -> var\_assign unload\_pv

Rule 30 statement -> statement statement

Rule 31 statement -> <empty>

Rule 32 var\_assign -> ID store\_oper = expr ;

Rule 33 returns -> RETURN expr store\_rquad

Rule 34 ifelse -> IF ( expr ) store\_gotof { ifelsecont } store\_endif

Rule 35 ifelse -> IF ( expr ) store\_gotof { ifelsecont } ELSE store\_goto { ifelsecont } store\_endif

Rule 36 ifelsecont -> <empty>

Rule 37 ifelsecont -> statement ifelsecont

Rule 38 output -> OUTPUT ( expr outex ) ;

Rule 39 outex -> <empty>

Rule 40 outex -> , expr outex

Rule 41 input -> INPUT ( ID store\_oper ) ;

Rule 42 loop -> WHILE store\_jump ( expr ) store\_gotof { loopcont } end\_loop

Rule 43 loopcont -> <empty>

Rule 44 loopcont -> statement loopcont

Rule 45 expr -> arexp arexpx

Rule 46 arexpx -> <empty>

Rule 47 arexpx -> arexpxop arexp

Rule 48 arexpxop -> OR

Rule 49 arexpxop -> EQEQ

Rule 50 arexpxop -> DIFF

Rule 51 arexpxop -> AND

Rule 52 arexpxop -> <

Rule 53 arexpxop -> LOETHAN

Rule 54 arexpxop -> >

Rule 55 arexpxop -> GOETHAN

Rule 56 arexp -> term arexpextra

Rule 57 arexp -> term

Rule 58 arexpextra -> <empty>

Rule 59 arexpextra -> - term arexpextra [precedence=left, level=1]

Rule 60 arexpextra -> + term arexpextra [precedence=left, level=1]

Rule 61 term -> factor termx

Rule 62 term -> factor

Rule 63 termx -> <empty>

Rule 64 termx -> / factor termx [precedence=left, level=2]

Rule 65 termx -> \* factor termx [precedence=left, level=2]

Rule 66 factor -> FLOAT store\_const

Rule 67 factor -> element

Rule 68 factor -> ( store\_op expr ) store\_op

Rule 69 element -> callfunc store\_oper

Rule 70 element -> compound store\_oper

Rule 71 element -> const store\_const

Rule 72 compound -> compoundx

Rule 73 compoundx -> ID store\_oper

Rule 74 const -> INT

Rule 75 const -> FALSE store\_const

Rule 76 const -> TRUE store\_const

Rule 77 callfunc -> ID verify\_func add\_fstack ( callfuncpar ver\_params ) end\_fstack store\_gosub

Rule 78 callfuncpar -> <empty>

Rule 79 callfuncpar -> expr store\_pquad callfuncparx

Rule 80 callfuncparx -> <empty>

Rule 81 callfuncparx -> , callfuncpar

Rule 82 datatype -> D\_BOOL

Rule 83 datatype -> D\_STRING

Rule 84 datatype -> D\_FLOAT

Rule 85 datatype -> D\_INT

Rule 86 check\_program -> <empty>

Rule 87 store\_init\_quad -> <empty>

Rule 88 store\_funcm -> <empty>

Rule 89 store\_funcv -> <empty>

Rule 90 store\_params -> <empty>

Rule 91 store\_local\_vars -> <empty>

Rule 92 store\_mainv -> <empty>

Rule 93 store\_op -> <empty>

Rule 94 store\_oper -> <empty>

Rule 95 store\_const -> <empty>

Rule 96 unload\_pv -> <empty>

Rule 97 store\_rquad -> <empty>

Rule 98 close\_func -> <empty>

Rule 99 store\_endif -> <empty>

Rule 100 store\_pquad -> <empty>

Rule 101 ver\_params -> <empty>

Rule 102 verify\_func -> <empty>

Rule 103 add\_fstack -> <empty>

Rule 104 end\_fstack -> <empty>

Rule 105 store\_gosub -> <empty>

Rule 106 store\_gotof -> <empty>

Rule 107 store\_goto -> <empty>

Rule 108 store\_jump -> <empty>

Rule 109 end\_loop -> <empty>

Descripción de Generación de Código Intermedio

Diagram, schematic

Description automatically generated

Se puede ver mejor en el diagrama dentro del folder del proyecto.

Esta es la descripción de las acciones semánticas de cada punto:

Store\_init\_quad: Se guarda el quad inicial tomando el tamaño de la lista de los quads y tomando el nombre de la función sacando la función del diccionario de funciones.

Store\_funcm: se toma el scope como main, se crea la función, se añade su quad inicial y se guarda en memoria además de en el diccionario de variables y de funciones.

Store\_funcv: Se guarda una función en el diccionario de variables y en el de funciones creando también una dirección para la misma.

Store\_params: Se guardan los parámetros que se han guardado en el stack de params.

Store\_mainv: Se pone la dirección del goto y se guardan las variables del main tomándolas del stack de local vars

Store\_op: se guarda la operación en el stack de operaciones

Store\_oper: Se verifica que la variable exista y si es así se guarda en el stack de operandos de los quads

Store\_const: Se guarda la constante tanto en el stack de operandos y se le da un address

Unload\_pv: se quita todo del polish vector

Store\_rquad: Se genera el return quad y se añade el operando de la función, y las operaciones “return” y “endfunc”

Store\_endif: Se termina en gotof

Store\_pquad: Se genera el quad de parámetros para añadirlos también al stack de operandos y la operación “params” al stack de operadores.

Ver\_params: Verifica que el numero de parámetros mandados sea el mismo al numero de los parámetros declarados

Verify\_func: Se verifica que la función esté en el diccionario de funciones

Store\_gosub: Se añade el operando de la función y el “gosub” a los operadores además de que se añade el return de la función

Store\_gotof: Se añade el operador “gotof”

Store\_goto: Se añade el operador “goto”

Store\_jump: Se añade el quad al stack de jumps

End\_loop: Se añade el quad “gotow” al stack de operadores

Descripción de Administración de Memoria:

Para la tabla de funciones se hizo una variable de funciones donde podemos tener los siguientes atributos:

Text

Description automatically generated

Y la tabla de funciones solamente es un diccionario de funciones donde se guarda el nombre de una función como key y el value sería el objeto función.

Para las variables se hizo lo mismo, se declaró una clase variable para almacenar los siguientes atributos:

Text

Description automatically generated

Y la tabla de variables igualmente es un diccionario de variables donde se guarda el nombre de una variable como key y el value seria el objeto variable. Se usaron clases porque usar diccionarios dentro de diccionarios sería más complicado a mi parecer.

Para las direcciones virtuales se hizo la delimitación de los datos como se hablo previamente ordenados entre ints, floats, strings y booleanos y cada uno definido en globales, locales y constantes.

Text

Description automatically generated

Cada uno con un contador que se va actualizando como se van insertando cada dato (que se hará en el parser en compilación), después el parser genera un diccionario con estos datos para poder trabajar con ello en ejecución, en la memoria de ejecución se declara un diccionario con arreglos con un numero de espacios basados en el numero de datos que el parser desarrollo en sus diccionarios.

A picture containing graphical user interface

Description automatically generated

Referencias:

https://ruslanspivak.com/lsbasi-part1/

Writing Compilers and Interpreters: A Software Engineering Approach

https://www.amazon.com/gp/product/193435645X/ref=as\_li\_tl?ie=UTF8&camp=1789&creative=9325&creativeASIN=193435645X&linkCode=as2&tag=russblo0b-20&linkId=MP4DCXDV6DJMEJBL

Language Implementation Patterns: Create Your Own Domain-Specific and General Programming Languages (Pragmatic Programmers)

Compilers: Principles, Techniques, and Tools (2nd Edition)