UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA
Escuela de Ingeniería Informática

Procesadores de Lenguaje:

WiC

Curso 2018 / 2019

invalid utf-8 byte sequence spanish characters listing latex

Realizado por:
Medina Medina, David Alberto
Brito Ramos, Christian
Hernández Delgado, Christopher
López González, Néstor

Índice

`WiC` (Winter is Coming) es un lenguaje de programación diseñado por estudiantes de la *ULPGC* con el objetivo de establecer el diseño y desarrollo de un compilador de lenguaje. El objetivo de esta experiencia es desarrollar un compilador del lenguaje que pase por las diferentes etapas de análisis léxico, sintáctico y semántico con el objetivo de generar un código intermedio en lenguaje ensamblador `IA-32`

1. Definición del lenguaje

En esta sección se describen todos los detalles que describen las características propias del lenguaje `WiC`.

1.1. Tipos de datos

Cualquier leguaje de programación necesita definir un conjunto de *tipos de datos*, esto es, la batería de valores y operaciones que puede adquirir una variable. Cada tipo de dato está definido en el lenguaje por un *literal* único que lo representa, lo que permite que cada tipo de dato tenga un representación física específica.
Los tipos de datos definidos en el lenguaje son los que figuran en el *cuadro 1*. Las características críticas de implementación que define a cada tipo son:

Entero Representa a todas y cada una de las variables enteras que sean declaradas en el lenguaje. Este tipo de dato presenta un tamaño de 4 bytes (32 bits) y permite representar números enteros con signo. El *complemento a 2* es el sistema elegido para definir el signo del número entero. Este dato se representa por el literal `int`. El rango de valores que puede tomar es

$$\left[-2^{N-1},\, 2^{N-1}-1\right] = \left[-2^{32-1},\, 2^{32-1}-1\right] = [-2147483648,\, 2147483647] \tag{1}$$

donde, N es el número de bits disponibles para representar el número entero (32 bits).

Coma flotante de simple precisión Este tipo de dato representa número reales en coma flotante de simple precisión con un tamaño de 4 bytes (32 bits) siguiendo el estándar *IEEE 754*. En la figura 1.1 puede observarse como esta representación binaria los bits se organizan en tres sectores principales:

- **Signo** (1 bit). Se trata de un sólo bit que define el signo del número: positivo (0) o negativo (1).
- **Exponente (8bits)**. Se trata de un número entero con signo de complemento a 2 ($[-128,\, 127]$)
- **Mantisa (23 bits)**. Conforma la fracción a la derecha de la coma binaria y un bit de encabezado implícito.

Este tipo de dato se representa con el literal `real`. Su rango de valores es de $\left[1{,}18 \cdot 10^{\check{}38},\ 3{,}4 \cdot 10^{38}\right]$.

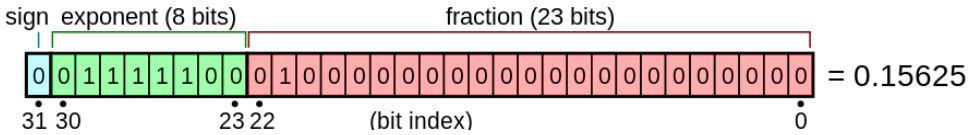

Figura 1: Representación binaria de número en coma flotante de simple precisión (*IEEE 754*)

Caracter Este tipo de dato es usado para representar caracteres con la codificación *ASCII*, es por este motivo que el tamaño de que ocupan las variables de tipo caracter son de 1 byte de tamaño, siendo `char` el literal que lo representa.

Booleano Se trata de un tipo de dato utilizado para representar representar valores booleanos. Su tamaño es de 1 bit, por lo que tan solo puede tomar dos valores: `1` (verdadero) ó `0` (falso). El literal que lo representa es `bool`.

Cuadro 1: Tipos de datos

Tipo	**Literal**	**Tamaño**	**Rango**
Entero	int	4 Bytes	$[-2147483648, 2147483647]$
Coma flotante de simple precisión	real	4 Byte	$[1{,}18 \cdot 10^{\check{}38}, 3{,}4 \cdot 10^{38}]$
Caracter	char	1 Byte	[`0x00 - 0xFF`]
Lógico	bool	1 Bit	[0, 1]

En el siguiente ejemplo se muestra cómo se declaran las variables con los literales de los tipos descritos anteriormente:

```
int entero      .. Esto es un entero
real flotante   .. Esto es un numero en coma flotante
char caracter   .. Esto es un caracter
bool booleano   .. Esto es un booleano
```

1.2. Colecciones de datos: `Arrays`

Las variables pueden ser agrupadas en colecciones de datos de una dimensión denominados `arrays`. En este lenguaje, cualquier tipo de dato puede formar parte de un `array`.
Para declarar un `array` del tipo que se desee, debe usar la gramática 2:

$$tipo[< int >]\ nombre\_variable \tag{2}$$

donde el *tipo* define el tipo de dato e $< int >$ un valor entero opcional que define el tamaño del `array`. Cuando no se indica este último valor entero, no se lleva a término la reserva en memoria del `array` declarado. En caso contrario, se reservará en memoria tantos bytes/bits como fueren necesarios para

generar una colección de tipos de datos del tamaño indicado por $< int >$. El número de bytes/bits a reservar está determinado por el tipo de dato y el tamaño del `array`,

$$Tamano_{memoria} = Tamano_{tipo\,dato} \times Tamano_{array} \tag{3}$$

Un `array` que ha sido declarado con anterioridad puede ser redefinido haciendo uso de la gramática 4,

$$tipo[\,]\ nombre\_variable = new\ tipo[< int >]\ \{expresion\_1, expresion\_2, ...\ \} \tag{4}$$

Esta notación es alternativa a la gramática 2, donde podremos inicializar el `array` a un conjunto de expresiones separados por comas encerrados dentro de los caracteres `{` y `}`. Estas expresiones son opcionales. El tamaño máximo del vector y, por tanto, de expresiones es el indicado por $< int >$ el cual es un número entero de caracter obligatorio.
Si se declara un `array` utilizando la gramática 2 o 4 sin expresiones en esta última, el array es inicializado en todas sus posiciones al valor `0` para tipos de datos enteros y reales. Para caracteres el valor por defecto es el caracter `nulo`. Y para tipos booleanos el valor por defecto es `FALSE`.
La gramática 5 es necesaria para acceder al valor de un elemento del `array` en una posición arbitraria del mismo,

$$nombre\_variable\ [< int >] \tag{5}$$

donde $< int >$ es un entero obligatorio que indica la posición del `array` a la que se desea acceder.
El `array` de caracteres constituyen los denominados `string`, los cuales requieren una atención especial ya que es posible cargar una variable con una secuencia de caracteres sin ser necesaria la declaración dada por la gramática 4. La gramática 6 define la instancia de un `string` con un conjunto de caracteres localizados entre los caracteres comillas doble,

$$nombre\_variable\ = " < char >< char > ..." \tag{6}$$

El siguiente listado muestra algunos ejemplos de uso de los `array` definidos en este lenguaje:

Listing 1: Ejemplo de uso de arrays

```
',
    En este bloque se declara un array de cada tipo de variable sin asignacion de espacio en
        memoria
.',

int [] entero
real [] real
char [] string
bool [] booleano


',
    En este bloque se declara un array de cada tipo con asignacion de espacio en memoria (de
        tamano 5 cada uno)
.',
```

```
int [5]  entero5
real [5]  real5
char [5]  string5
bool [5]  booleano5


'.
   En este bloque se declara un array de cada tipo y se instancia cada uno de los valores de los
       elementos que lo componen
.'

int [2]  entero2 = { 0, 1}
real [2]  real2 = {0.0, 1.0}
char []  string2_ab = "ab"  .. Declaracion con asignacion alternativo al tipo string
char [2]  string2_cd = {'c','d'}
bool [4]  booleano2 = {true, false, true, true}
```

1.3. Palabras reservadas y cambios de contexto

Este lenguaje utiliza una serie de palabras reservadas que se utilizan para desempeñar las funciones aquí descritas:

`continue` La sentencia de continue es de tipo de control de bucles. Dentro de la iteracion en un bucle, de cualquiera de los tipos (while, do-while, for), el uso de esta sentencia rompe la iteracion de dicho bucle. Provocando que se ejecute la siguiente iteracion de dicho bucle, ignorando las sentencias posteriores a çontinue".

`break` Dentro de la iteracion en un bucle, de cualquiera de los tipos (while, do-while, for), el uso de esta sentencia rompe la iteracion de dicho bucle.

`return` Palabra empleada para retornar el resultado de un método o función, además de interrumpir la ejecución del mismo.

`void` Es utilizado para indicar que una función no debe devolver ningún valor.

`fun` Palabra especial utilizada al principio de cada declaración o definición de funciones.

Los **cambios de contexto** de bloques de instrucciones son aplicables cuando se hace uso de un caracter `0x09` (`TAB`) o de 4 espacios consecutivos (`0x20`).

1.4. Comentarios

Aquí va el texto. Poner siempre un código de ejemplo.

- **'Comentario de línea'** . Para indicar un comentario de línea, basta con comenzar con '..', siendo el resultado final '.. TEXTO'.

Listing 2: Ejemplo de comentario de línea

```
int a = 7
.. Variable que almacena una suma.
int suma = 0
suma = 30 + a
```

- **'Comentario de bloque'** . Los comentarios de varias líneas se establecen utilizando ',.' y '.,' quedando como resultado ',. VARIAS LÍNEAS DE TEXTO .,'.

Listing 3: Ejemplo de comentario de bloque

```
int a = 7
int suma = 0
suma = 30 + a
,.
En este punto, la variable suma toma el
valor de 37.
.,
```

1.5. Tipos de operadores

El conjunto de posibles operadores aplicables a nuestro lenguaje pueden clasificarse en:

- Operadores aritméticos
- Operadores lógicos
- Operadores bit a bit
- Operadores de array

1.5.1. Operadores aritméticos

Estos operadores son necesarios para realizar operaciones matemáticas sencillas. La descripción detallada de cada uno de estos operadores queda reflejado en la tabla 2.

Cuadro 2: Operadores aritméticos

Operación	**Tipo operación**	**Lexema**	**Estructura**	**Tipo expresión**	**Tipo evaluado** [1]
Suma	Binaria	`+`	`{Expr} + {Expr}`	`int/real`	`int/real`
Resta	Binaria	`-`	`{Expr} - {Expr}`	`int/real`	`int/real`
Multiplicación	Binaria	`*`	`{Expr} * {Expr}`	`int/real`	`int/real`
División	Binaria	`/`	`{Expr} / {Expr}`	`int/real`	`int/real`
Módulo	Binaria	`%`	`{Expr} % {Expr}`	`int/real`	`int/real`
Potencia	Binaria	`**`	`{Expr} ** {Expr}`	`int/real`	`int/real`
Raíz	Binaria	`#`	`{Expr} # {Expr}`	`int/real`	`int/real`
Pre-incremento [2]	Unaria	`++`	`++{Expr}`	`int/real`	`int/real`
Pre-decremento [2]	Unaria	`--`	`--{Expr}`	`int/real`	`int/real`
Post-incremento [3]	Unaria	`++`	`{Expr}++`	`int/real`	`int/real`
Post-decremento [3]	Unaria	`--`	`{Expr}--`	`int/real`	`int/real`

[1] Todas las operaciones binarias permiten entremezclar el uso de un valor entero y otro real al mismo tiempo aunque esto conllevaría el *casteo* del resultado de la operación a un valor de tipo real.
[2] El valor de la variable es incrementado/decrementado una unidad previa asignación del mismo.
[3] El valor de la variable es incrementada/decrementada una unidad después de asignar el valor del misma.

Listing 4: Ejemplo de uso los operadores artiméticos

```
int a = 2
int b = 4

int c = 2#4          .. c = 2

int d
d = c++              .. d = 2, c = 3

int e
e = ++c              .. e = 4, c = 4
```

1.5.2. Operadores lógicos

Definimos los símbolos que identificarán a los operadores lógicos de forma análoga a los utilizados por gran parte de otros lenguajes de programación, separando cada operador 2 expresiones a comparar por los operadores lógicos a utilizar, con la excepción del propio *NOT*, que podrá utilizarse directamente para determinar si el valor devuelto por una expresión de tipo `booleano` es directamente verdadero o no.
Como dato adicional, también consideramos la posibilidad de hacer uso de comparadores mediante dichos operadores lógicos n-arios, comparando pares de expresiones entre sí de forma sucesiva siguiendo

el orden de lectura de izquierda a derecha (que a efectos prácticos es lo mismo que hacer sucesivos *AND* pero de forma más breve) y la posibilidad de prioridades para ciertos operadores (como el *AND*).

En relación a las expresiones de comparación, consideramos hacer que de momento solo se puedan utilizar entre tipos de datos del mismo tipo o similar (por ejemplo, entre enteros y reales podría hacerse). Dejamos la tabla de operadores como sigue:

Cuadro 3: Operadores lógicos

Operación	**Tipo operación**	**Lexema**	**Estructura**	**Tipo expresión**[1]	**Tipo evaluado**
`AND`	Binaria	`&&`	`{Expr} && {Expr}`	`int/real/bool`	`bool`
`OR`	Binaria	`||`	`{Expr} || {Expr}`	`int/real/bool`	`bool`
`NOT`	Unaria	`!`	`!{Expr}`	bool	`bool`
Igual	Binaria	`==`	`{Expr} == {Expr}`	`int/real/bool`	`bool`
No Igual	Binaria	`!=`	`{Expr} != {Expr}`	`int/real/bool`	`bool`
Mayor	Binaria	`>`	`{Expr} >{Expr}`	`int/real/bool`	`bool`
Mayor o igual	Binaria	`>=`	`{Expr} >= {Expr}`	`int/real/bool`	`bool`
Menor	Binaria	`<`	`{Expr} <{Expr}`	`int/real/bool`	`bool`
Menor o igual	Binaria	`<=`	`{Expr} <= {Expr}`	`int/real/bool`	`bool`

[1] El tipo de dato utilizado como expresión deben ser del mismo tipo para operadores lógicos binarios.

Listing 5: Ejemplo de uso los operadores lógicos

```
int a = 7
int suma = 0
5 <= a < 10 ?:
suma = 4 * a

suma > 20 > a && 2*suma < a
suma++
```

1.5.3. Operadores bit a bit

Estos operadores permiten manipular los valores de una expresión a nivel de bits. El cuadro 4 refleja la descripción detallada de cada uno de estos operadores.

Cuadro 4: Operadores bit a bit

Operación	Tipo operación	Lexema	Estructura	Tipo expresión[1]	Tipo evaluado[2]
AND	Binaria	&	{Expr} & {Expr}	Cualquiera	Cualquiera
OR	Binaria	|	{Expr} | {Expr}	Cualquiera	Cualquiera
XOR	Binaria	^	{Expr} ^ {Expr}	Cualquiera	Cualquiera
Desp. izq.	Unaria	«	{Expr} « {Expr}[3]	Cualquiera	Cualquiera
Desp. dcha.	Unaria	»	{Expr} » {Expr}[3]	Cualquiera	Cualquiera

[1] El segundo operando debe ser siempre de tipo `int`.
[2] El tipo de dato asignado a la variable debe ser del mismo tipo que el del primer operando.
[3] Rota los bits de la expresión de la izquierda tantas unidades a la izquierda/derecha como indique la expresión de la derecha.

Listing 6: Ejemplo de uso los operadores bit a bit

```
char c = ' '          .. c ==> Caracter ASCII 0x20
int  i = 2

char rc = c<<2        .. rc = '@' ==> Caracter ASCII 0x40
int  ri = i<<2        .. ri = 8
```

1.5.4. Operadores de array

Los operadores de arrays nos permiten realizar operaciones con arrays, tratándolos como a subconjuntos. Los operadores a definir son los siguientes:

Unión La unión de dos arrays consiste en crear un nuevo array formado por los elementos del primer array y los del segundo. La unión no permite elementos repetidos.

Diferencia La diferencia de un array con otro consiste en crear un nuevo array con todos los elementos del primer array, descartando aquellos que aparecen en el segundo.

Intersección La intersección de dos arrays consiste en crear un nuevo array formado por todos los elementos comunes entre el primer y el segundo array.

Concatenación Nuevo array formado por los elementos del primero seguido de los del segundo. La concatenación permite crear el nuevo array con elementos repetidos.

Cuadro 5: Operadores de array

Operación	Tipo operación	Lexema	Estructura	Tipo expresión	Tipo evaluado
Unión	Binaria	`U`	`{Expr} U {Expr}`	Cualquiera	Cualquiera
Diferencia	Binaria	`D`	`{Expr} D {Expr}`	Cualquiera	Cualquiera
Intersección	Binaria	`I`	`{Expr} I {Expr}`	Cualquiera	Cualquiera
Concatenación	Binaria	`+`	`{Expr} + {Expr}`	Cualquiera	Cualquiera

[1] Rota los bits de la expresión de la izquierda tantas unidades a la izquierda/derecha como indique la expresión de la derecha.

A continuación mostramos un ejemplo de uso de los operadores de array para un caso trivial meramente ejemplificativo:

Listing 7: Ejemplo de uso de operadores de array

```
int [6]  a = {0,1,2,3,4,5}
int [7]  b = {3,4,5,6,7,8,9}

int []  c = a U b                  .. c =  {0,1,2,3,4,5,6,7,8,9}
int []  d = a D b                  .. d = {0,1,2}
int []  e = a I b                  .. e = {3,4,5}
int []  f = (a U b) D (a I b)          .. f = {0,1,2,6,7,8,9}
int []  g = d + f + a              .. g =  {0,1,2,0,1,2,6,7,8,9,0,1,2,3,4,5}
```

1.6. Estructuras de control

El conjunto de instrucciones del lenguaje no tienen por qué ejecutarse en una secuencia lineal sino que es posible que el programador establezca un cierto control en esa secuencia de ejecución. Las diferentes estructuras de control que se definen en nuestro lenguaje se clasifican en los siguiente grupos:

- Sentencias `if-ifelse-else`
- Bucle `for-forelse-else`
- Bucle `while-whileelse-else`

1.6.1. Sentencias `if-ifelse-else`

La estructura de control if-ifelse-else sentencia condicional que está compuesta de los siguientes bloques:

Bloque `if` Este bloque está identificado por la siguiente estructura:

```
{Expr} ?:
{Instr}
...
```

En caso de ser verdadera, se ejecutan las instrucciones de este bloque. Sólo puede existir un único bloque `if` al principio de una sentencia `if-ifelse-else`.

Bloque `ifelse` Si no se cumple la condición del bloque anterior, se comprueba si la condición de este bloque se cumple para posteriormente ejecutar el conjunto de instrucciones que se encuentran en su contexto. En caso contrario, pasa a ejecutarse el siguiente bloque `elseif` La estructura de este bloque es la siguiente:

```
. {Expr} ?:
{Instr}
...
```

Bloque `else` Si ninguna de la expresiones de los bloques anteriores cumple la condición, se ejecutarán las instrucciones contenidas en este bloque. La estructura de este bloque es la siguiente:

```
.?:
{Instr}
...
```

Sólo puede existir un único bloque `else` al final de una sentencia `if-ifelse-else`.

Listing 8: Ejemplo de uso de la sentencia `if-ifelse-else`

```
int a = 25
int b = 0

a > 20 ?:
b = a − 10
. a < 10 ?:
b = a + 5
.?:
b = −1
```

1.6.2. Bucle `for-forelse-else`

Este tipo de bucle permiten modificar el valor de una variable mientras se realiza la evaluación lógica de uno de sus bloques. El bucle for-forelse-else se estructura en los siguientes bloques:

Bloque `for` Este bloque presenta la siguiente estructura:

{Expr} ?? {Expr}:
{Instr}
...

La primera expresión no exige que figure el parámetro utilizado en la iteración. Si, por ejemplo, se definen otras formas de salida dentro del contexto del bucle, y si se cumple la condición evaluada se ejecutarán las instrucciones dentro del contexto en caso de que el conjunto de condiciones del bucle se cumpla. Además, es legal no indicar una expresión lógica a evaluar, en cuyo caso, el resultado de la evaluación siempre será `true`, iterándose indefinidamente las instrucciones de este contexto.

La segunda expresión se usará comúnmente como modificador del valor de iteración empleado en la condición.

Bloque `forelse` Se ha considerado hacer que exista el bloque `for-else`, que actuaría de tal manera que si no se cumple inicialmente la condición del primer bucle, comprobará la expresión del siguiente, y si su condición se cumple, estará evaluando repetidamente las instrucciones de este último (sin volver a comparar con las condiciones del bucle anterior al `else`). De esta manera, podemos hacer una estructura combinada de lo que en otros casos serían `if-for-else-for`, de forma directa con solo un `for-else`. Su estructura es la siguiente:

. {Expr} ?? {Expr}:
{Instr}
...

Bloque `else` De forma análoga a los `else` de las expresiones condicionales, la representación para este bloque es la siguiente:

.??:
{Instr}
...

Este bloque es siempre terminal al bucle `for-forelse-else` y siempre iterará por el conjunto de instrucciones de su contexto (bucle infinito) siempre y cuando las expresiones lógicas de los bloques anteriores evalúen a `false`.

Listing 9: Ejemplode uso del bucle for-forelse-else

```
int a = 7
```

```
int b = 9
int suma = 0
int contador = 0

a > contador ?? contador++:
suma = suma + a * contador + b
. b > contador ?? contador++:
suma = suma + b*contador + a
.??:
suma = a + b
.. Como 'a' es mayor que 'contador' inicialmente, solo se ejecutan instrucciones del primer
   bucle
```

1.6.3. Bucle `while-whileelse-else`

Este bucle permite ejecutar un bloque de instrucciones de manera iterativa siempre y cuando la condición de la expresión sea `true`. En caso contrario, se irá probando las diferentes condiciones de cada bloque del bucle. Este bucle se puede estructura en los siguientes bloques:

Bloque `while` Este bloque está identificado por la siguiente estructura:

```
{Expr} ??:
{Instr}
...
```

Mientras la expresión del bloque siga sienda `true`, se ejecuta el bloque de instrucciones del bucle. Si la expresión es `false`, se evalúa la expresión lógica del siguiente bloque. Sólo puede existir un único bloque `while` al principio de un bucle `while-whileelse-else`.

Bloque `whileelse` Si no se cumple la condición del bloque anterior, se comprueba si la expresión lógica de este bloque es `true`. Si es así, se itera por el conjunto de instrucciones que se encuentran en su contexto hasta que la expresión sea `false`. Si la evaluación de la expresión lógca es `false`, se pasa a ejecutar el siguiente bloque `whileelse` o el bloque terminal `else`. La estructura de este bloque es la siguiente:

```
. {Expr} ??:
{Instr}
...
```

Bloque `else` Si ninguna de la expresiones de los bloques anteriores cumple la condición, se iterarán las instrucciones contenidas en este bloque indefinidamente, generándose un bucle infinito termnial. La estructura de este bloque es la siguiente:

```
.??:
{Instr}
...
```

Sólo puede existir un único bloque `else` al final de un bucle `while-ifelse-else`.

En cualquier caso, la expresión lógica es opcional. Si no se indica una existe una expresión al principio del bucle, se itera por el conjunto de instrucciones de ese bloque de manera indefinida.

1.7. Declaración, definición y uso de funciones

Las funciones componen una estructura fundamental en cualquier lenguaje de programación que son útiles para estructurar funcionalmente el código en diferentes contextos, aportando una mayor expresividad del lenguaje.
Una función puede ser declarada sin especificar el código que se ejecutará dentro de su contexto. La forma general de declaración puede observarse en el listado 10.

Listing 10: Declaración de funciones

```
fun {tipo} {nombre_funcion} ({parametro1}, [{parametro2}], ...)
```

La definición de cualquier función sigue la estructura del listado 11. El cuerpo de la función debe comenzar por el caracter de cambio de contexto (caracter `0x09` o cuatro caracteres `0x20` seguidos).

Listing 11: Definición de funciones

```
fun {tipo} {nombre_funcion} ({parametro1}, [{parametro2}], ...):
    {instruccion1}
    [{instruccion2}]
     ...
```

El listado 12 ilustra la estructura general de llamada a una función.

Listing 12: Declaracion de funciones

```
{nombre_funcion}({parametro1}, [{parametro2}], ...)
```

Es importante recalcar que la referencia en memoria de un array es leída cuando este es usado como parámetro de una función. Si un array es el valor de retorno de una función, se devuelve su referencia en memoria.

1.8. Funciones primitivas

En nuestro lenguaje, Van a existir predefinidas un conjunto de funciones primitivas que facilitarán al programador su trabajo en el desarrollo del programa. Las funciones están divididas en las siguientes categorías:

Interacción entrada-salida Las funciones básicas de E/S son:

`bool print(T salida)` Esta función permite mostrar por pantalla el valor de `salida`, cuyo tipo `T` puede ser `int (%d), real(%r), bool(%b), char(%c)` o una string (`%s`). Si la operación termina correctamente la primitiva devolverá `true`, en caso contrario, devuelve `false`.

`T scan(T entrada)` Esta rutina permite leer el valor pasado por teclado una vez se pulse `ENTER` (`0x0D`). Si la operación termina satisfactoriamente la primitiva devolverá un valor de tipo `T`, y en caso de que no se especifique, devuelve el valor en formato `string`. Por defecto se lee como una String, pero se le puede indicar que tipo de dato se quiere leer: %d , %r , %b , %c o %s.

Listing 13: Ejemplo de uso de la función `print()` y `scan()`

```
int number
char[] name
print("Introduzca el primer número: ")
number = scan(%d)
print("¿Cuál es su nombre? ")
name = scan(%s)
print("El usuario %s ha introducido el número %d", name, number)
```

Funciones para arrays numéricos Las funciones para arrays numéricos permiten operaciones que impliquen el uso de arrays. El tipo `T` de las siguientes funciones indica que el tipo del parámetro de entrada o el tipo de retorno de la función puede ser `int` , `real` , `bool` o `string`:

`int length(array)` Devuelve el número de elementos del array.

`T[] delete(T[] array, int position)` Elimina el elemento de la posición `position` del array y devuelve la referencia al array resultante. Los demás elementos a partir de la posición especificada serán desplazados una posición a la izquierda.

`bool equal(T[] array1, T[] array2)` Devuelve `true` si los dos arrays son iguales en tamaño y los elementos están dispuestos en el mismo orden. En caso contrario, devuelve `false`.

Listing 14: Ejemplo de uso de funciones primitivas para arrays numéricos

```
int[] v1 = {2,5,9,5}
int[] v2 = {5,4,8,9,3}
int tam = length(v1)          .. tam = 4
delete(v1,1)                  .. v1 = {5,9,5}
boolean b1 = equal(v2,v1)     .. b1 = false
boolean b2 = equal(v1,v1)     .. b2 = true
```

1.9. Código ejemplo

Por último, mostraremos varios ejemplos de programas en WiC con la finalidad de abarcar las subsecciones anteriores y tener una visión general de como funciona el lenguaje de programación.

El primer código de ejemplo consiste en mostrar por pantalla una suscesión de fibonacci que termina cuando un valor de la serie alcanza el valor de un dato introducido por teclado.

Listing 15: Ejemplo de la función Fibonacci con parametros introducidos.

```
.. Variable global que contiene el resultado
global int a = 0

.. Funcion que actualiza los valores de a y b en funcion de sus valores anteriores
fun void fibo()
static int b = 1
int aux = b
b = a + b
a = aux

.. Leer de teclado un entero
n = scan(int)

.. Bucle que finaliza cuando a alcanza el valor introducido por teclado.
( a < n ) ??:
print(a)
print(', ')
fibo()
print( crlf )
```

Listing 16: Salida en consola para " n -> 6 ".

```
0, 1, 1, 2, 3, 5
```

A continuación, mostramos el código de un programa que invierte el contenido de un array. Además, añadimos la condición de que en las posiciones impares, se sustituye el valor por un 0.

Listing 17: Cambiar el orden de los elementos en un Array con condición.

```
.. Inicializamos el vector que se va a invertir
int [10] miVector = {0,1,2,3,4,5,6,7,8,9}

.. Inicializamos el vector que contendrá el resultado, reservando el tamaño en memoria
```

```
int [10] resultado

.. Variable que contendrá el tamaño del vector a invertir
int tamaño = length(miVector)

.. Variable que emplearemos de contador para el bucle
int contador = 0

.. Recorremos las posiciones de los vectores
contador <= tamaño ?? contador++:
.. Si la posición actual es 'par' ...
contador\ %2 == 0 ?:
resultado [contador] = miVector[tamaño-contador-1]
.. Si no es 'par'
.?:
miVector [contador] = 0


.,
Procedemos a imprimir el resultado con un bucle.
Se precisa resetear el valor del contador.
,.
contador = 0
print(''RESULTADO -> [ '')

contador < tamaño-1 ?? contador++:
print(resultado[contador])
print(' ', ' ')

print(resultado[tamaño-1])
print(' ']'')
```

2. Análisis léxico

En esta sección se describen todos los detalles que describen al analizador léxico del lenguaje `WiC`.

2.1. Introducción

El analizador léxico es la parte del compilador que se encarga de leer el flujo de caracteres que componen el programa fuente y agruparlos en tokens, que son agrupaciones de secuencias que guardan cierto significado; produciendo, para cada lexema, un token con la estructura $< nombre-token, valor-atributo >$. Cada token pasará a la fase del análisis sintáctico, a partir de la cuál se generará la gramática y el *Árbol Sintáctico Abstracto* (AST) asociado.

En cada token, el primer componente (nombre-token) se utiliza para el análisis sintáctico, y el segundo (valor-atributo), el cual es opcional, apunta a una entrada de la tabla de símbolos con el valor de dicho token.

Listing 18: Ejemplo de conversión de un conjunto de caracteres de un programa WiC en tokens

```
int doble = total * 2

(int)    ->  INT_TYPE
(doble) -> ID
=     ->  ASSIGN
(total) -> ID
*     ->  PRODUCT
2     -> INT_VAL
```

2.2. Tabla de tokens

Token	Valor-Atributo	Expresión-Regular
NEW	NO	(new)
CONTINUE	NO	$(continue)$
BREAK	NO	$(break)$
RETURN	NO	$(return)$
VOID	NO	$(void)$
FUN	NO	(fun)
GLOBAL	NO	$(global)$
STATIC	NO	$(static)$
INT_VAL	SI	$[0-9]+$
REAL_VAL	SI	$([0-9]+\.)[0-9]*\|\.[0-9]+$
BOOL_VAR	SI	$(true\|false)$
CHAR_VAL	SI	<CHAR_STATE>$[\hat{}']?\|(\\[nrtvb'"])$
STRING_VAL	SI	<STRING_STATE>$[\hat{}"]\|(\\[nrtvb'"])$
CHAR_QUOTE	NO	<CHAR_STATE>\' ó \'
STRING_QUOTE	NO	<STRING_STATE>\" ó \"
INT_TYPE	NO	(int)
REAL_TYPE	NO	$(real)$
BOOL_TYPE	NO	$(bool)$
CHAR_TYPE	NO	$(char)$
IF_CLAUSE	NO	(\?){1}
FOR_WHILE_CLAUSE	NO	(\?){2}
HEADER_END	NO	:
ELSE_IF_FOR_WHILE_CLAUSE	NO	(\.){1}
OPEN_CONTEXT_TAG	NO	<CHECK_LVL>[ ]*
CLOSE_CONTEXT_TAG	NO	<CHECK_LVL>[ ]*
CLOSE_CONTEXT_TAG	NO	<CHECK_LVL>«EOF»
ID	SI	$[a-zA-Z]([a-zA-Z]\|[0-9])+$
ASSIGN	NO	=
SUM	NO	\+
SUBSTRACT	NO	−
PRODUCT	NO	*
DIVIDE	NO	/
MODULUS	NO	%
POWER	NO	\ * *
RADICAL	NO	#
INCREMENT	NO	++
DECREMENT	NO	−−

Nombre-Token	**Valor-Atributo**	**Expresión-Regular**
AND_BIT	NO	&
OR_BIT	NO	|
XOR_BIT	NO	^
LEFT_SHIFT	NO	<<
RIGHT_SHIFT	NO	>>
AND	NO	&&
OR	NO	||
NOT	NO	!
EQUALS	NO	==
NOT_EQUALS	NO	!=
GREATER	NO	<
GREATER_EQUALS	NO	>=
LESS	NO	<
LESS_EQUALS	NO	<=
UNION	NO	U
DIFFERENCE	NO	D
INTERSECTION	NO	I
SQUARE_BRACKET_OPEN	NO	\[
SQUARE_BRACKET_CLOSE	NO	\]
CURLY_BRACKET_OPEN	NO	\{
CURLY_BRACKET_CLOSE	NO	\}
ELEM_SEPARATOR	NO	,
PARETHESES_OPEN	NO	\(
PARETHESES_CLOSE	NO	\)
END_OF_INSTR	NO	$\backslash n$

2.3. Estados

El estado de un analizador léxico es la situación en la que se encuentra a la hora de interpretar los caracteres del programa fuente. Los estados posibles del analizador léxico son:

CHECK_LVL Este estado viene por defecto. En él, se comprueba el nivel en el que se encuentra el programa. Se llega a este estado al principio del programa o cuando se detecta un caracter de salto de línea. En todo estado de CHECK_LVL, se comprueba el número de tabuladores (tomando el valor de un tabulador equivalente a cuatro espacios); con cada tabulador encontrado, se devuelve un token OPEN_CONTEXT_TAG si accedemos a un nivel estrictamente superior al que ya estábamos, y se devuelve CLOSE_CONTEXT_TAG si accedemos a un nivel estrictamente inferior al que ya estábamos. En caso de que ambos niveles sean exactamente iguales, no se devuelve ningún token, interpretando que no hay cambio de nivel.

INITIAL En este estado, los caracteres son interpretados normalmente y se generan los tokens correspondientes.

COMMENT_BLOCK Este estado corresponde a los comentarios de bloque. Se llega a este estado cuando se detecta la expresión regular ",¨. A partir de este estado, todos los caracteres serán considerados como parte de un comentario y no se tendrán en cuenta en la generación de tokens. Cuando se detecta la expresión regular ";", se vuelve al estado INITIAL.

COMMENT_LINE Este estado corresponde a los comentarios de línea. Se llega a este estado cuando se detecta la expresión regular "\.\.". A partir de este estado, todos los caracteres serán considerados como parte de un comentario y no se tendrán en cuenta en la generación de tokens. Cuando se detecta , se vuelve al estado INITIAL.

CHAR_STATE Este estado corresponde a los caracteres. Se llega a este estado cuando se detecta el caracter comilla simple ("´). En este estado, solo es permitido que aparezca como máximo un caracter o dos si el primero es la barra invertida (’́). Cuando se detecta otro caracter comilla simple, se vuelve al estado INITIAL.

STRING_STATE Este estado corresponde a las strings o cadenas de caracteres. Se llega a este estado cuando se detecta un caracter comillas dobles ("¨). A partir de este estado, todos los caracteres consecuentes formarán parte de una string. Cuando se detecta otro caracter comillas dobles, se vuelve al estado INITIAL.

3. Análisis sintáctico

3.1. El analizador sintáctico

El analizador sintáctico o parser constituye una representación intermedia del análisis de un programa. Su función es construir un árbol de análisis a partir de los componentes léxicos que recibe, aplicando las producciones de la gramática con el objetivo de comprobar la correctitud sintáctica del programa. Su salida es una representación de la cadena de componentes léxicos en forma de árbol sintáctico.

Dicho de otro modo: el analizador sintáctico comprueba que el orden en que el analizador léxico le va entregando es válido, y para ello verifica que la cadena pueda ser generada por la gramática del lenguaje fuente.

Otra de las funciones del analizador sintáctico es la de informar acerca de los errores en la sintaxis; recuperándose de los mismos, siempre que sea posible, para continuar procesando la entrada.

Para comprobar si una cadena pertenece al lenguaje generado por una gramática, los analizadores sintácticos construyen una representación en forma de árbol de 2 posibles maneras:

Analizadores sintácticos descendentes (Top-down) Construyen el árbol sintáctico desde la raíz hasta las hojas, partiendo del símbolo inicial de la gramática (axioma) y expandiendo las producciones hasta llegar a la cadena de entrada.

Analizadores sintácticos ascendentes (Bottom-up) Construyen el árbol sintáctico comenzando por las hojas, partiendo de los terminales de entrada y, mediante reducciones, llegan hasta el símbolo inicial. En este apartado, nos centraremos en los analizadores sintácticos ascendentes.

3.2. Ejemplo de análisis sintáctico

Esta instrucción genera los siguientes tokens:

---

int i = 0 -> INT_TYPE ID ASSIGN INT_VAL END_OF_INSTR

En este caso en concreto, se harán uso de las siguientes gramáticas:

---

input := instr END_OF_INSTR input | instr END_OF_INSTR
instr := data_init ID ASSIGN expr
data_init := data_type
data_type := INT_TYPE
expr := INT_VAL

El analizador, cuando llega un token de entrada, puede hacer dos cosas:

Si el token es una parte derecha de una regla (o completa una parte derecha en construcción), se puede deshacer la regla. Esta acción es llamada reducción.

Si no se puede completar una parte derecha, se acumula a lo que se está construyendo. Esta acción es llamada desplazamiento.

En este caso, el analizador sintáctico realiza un análisis ascendente de la siguiente forma:

Cadena de entrada: INT_TYPE ID ASSIGN INT_VAL END_OF_INSTR

Reducir data_type := INT_TYPE data_type ID ASSIGN INT_VAL END_OF_INSTR

Reducir data_init := data_type data_init ID ASSIGN INT_VAL END_OF_INSTR

Desplazar data_init, ID, ASSIGN

Reducir expr := INT_VAL data_init ID ASSIGN expr END_OF_INSTR

Reducir instr := data_init ID ASSIGN expr instr END_OF_INSTR

Reducir input := instr END_OF_INSTR input

4. Descenso recursivo

4.1. Introducción

El análisis sintáctico por descenso recursivo consiste en realizar un conjunto de procedimientos, uno para cada nodo no terminal. El algoritmo para un nodo terminal general tiene la siguiente estructura:

```
void A () {
    elegir una producción A, A -> X1X2...Xk;
    for (i = 1 a k) {
        if (X es un nodo no terminal) {
            llamar al procedimiento Xi();
        } else if (X es igual al sÃmbolo de entrada actual a) {
            avanzar la entrada hasta el siguiente sÃmbolo;
        } else {
            /*No ha ocurrido ningún error*/
        }
    }
}
```

Para volver recursivo este algoritmo, se elige más de una opción de producción. Para poder probar otra producción A, debemos establecer el apuntador de entrada a la posición inicial.

```
void A () {
    elegir una producción A|B, A -> X1X2...Xk, B -> Y1Y2...Yk;
    for (i = 1 a k) {
        if (X es un nodo no terminal) {
            llamar al procedimiento Xi();
        } else if (X es igual al sÃmbolo de entrada actual a) {
            avanzar la entrada hasta el siguiente sÃmbolo;
        } else {
            return 1;
        }
    }
}
```

NOTA: el fallo del else (return 1) no sugiere que el aná lisis finaliza ahÃ.

4.2. Ejemplo de análisis sintáctico por descenso recursivo

Esta instrucción genera los siguientes tokens:

---

5 > i ?? i++: -> INT_VAL, GREATER, ID, FOR_WHILE_CLAUSE, ID, INCREMENT, HEADER_END, END_OF_INSTR

En este caso en concreto, se harán uso de las siguientes gramáticas:

---

input := instr END_OF_INSTR input | instr END_OF_INSTR
instr := for_instr
for_instr := expr FOR_WHILE_CLAUSE expr HEADER_END END_OF_INSTR input
expr := expr SUM term | ID INCREMENT | term
term := INT_VAL

4.3. Gramática para análisis sintáctico por descenso recursivo

Para el descenso recursivo, hemos decidido utilizar el ámbito de la instrucción while para mostrar un ejemplo de cómo realizar el análisis sintáctico por descenso recursivo, dejando la gramática siguiente:

```
while_instr :=    expr FOR_WHILE_CLAUSE HEADER_END END_OF_INSTR
    OPEN_CONTEXT_TAG input CLOSE_CONTEXT_TAG

input:    instr  input |
      /* empty */

instr :   data_value ID END_OF_INSTR |
      data_value ID ASSIGN term END_OF_INSTR |
      ID ASSIGN term END_OF_INSTR

term: ID
      ID SUM term |
      ID SUBSTRACT term |
      ID PRODUCT term |
      ID DIVIDE term |
      data_value
      data_value SUM term |
      data_value SUBSTRACT term |
      data_value PRODUCT term |
      data_value DIVIDE term

expr: term |
      term LESS term |
      term GREATER term |
      term EQUALS term

data_value: INT_VAL |
          REAL_VAL |
          BOOL_VAL
```

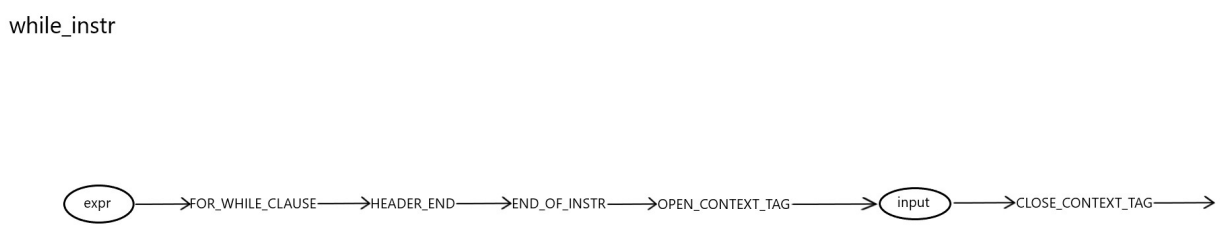

Diagrama sintáctico para while_instr

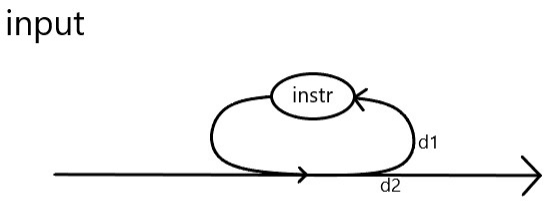

Diagrama sintáctico para input

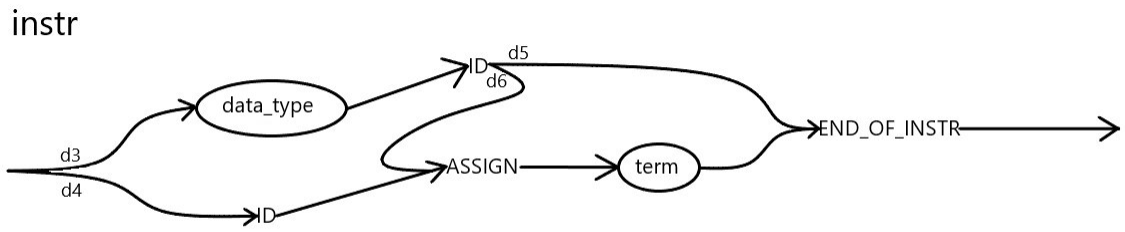

Diagrama sintáctico para instr

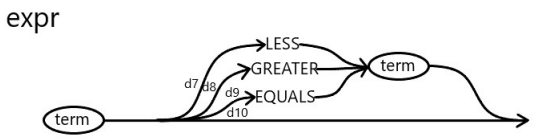

Diagrama sintáctico para expr

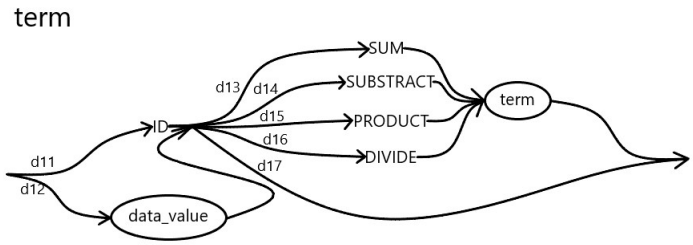

Diagrama sintáctico para term

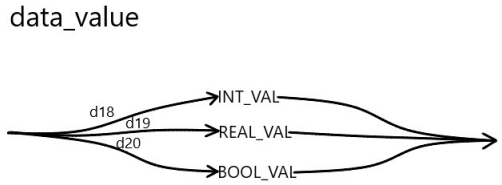

Diagrama sintáctico para data_value

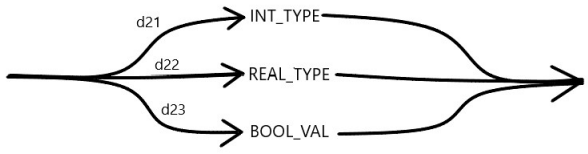

Diagrama sintáctico para data_type

no_terminal	**primeros**	**siguientes**
while_instr	ID, INT_VAL, REAL_VAL, BOOL_VAL	$
input	ID, INT_VAL, REAL_VAL, BOOL_VAL, empty	CLOSE_CONTEXT_TAG
instr	ID, INT_VAL, REAL_VAL, BOOL_VAL	ID
		INT_VAL
		REAL_VAL
		BOOL_VAL
		CLOSE_CONTEXT_TAG
expr	ID, INT_VAL, REAL_VAL, BOOL_VAL	FOR_WHILE_CLAUSE
term	ID, INT_VAL, REAL_VAL, BOOL_VAL	LESS
		GREATER
		EQUALS
		ID
		INT_VAL
		REAL_VAL
		BOOL_VAL
		END_OF_INSTR
data_value	INT_VAL, REAL_VAL, BOOL_VAL	LESS
		GREATER
		EQUALS
		ID
		INT_VAL
		REAL_VAL
		BOOL_VAL
		END_OF_INSTR
		SUM
		SUBSTRACT
		PRODUCT
		DIVIDE
data_type	INT_TYPE, REAL_TYPE, BOOL_TYPE	ID

Conjuntos de primeros y siguientes de cada símbolo no terminal

conjunto_director	grupo	siguientes
d1	1	ID, INT_TYPE, REAL_TYPE, BOOL_TYPE
d2	1	CLOSE_CONTEXT_TAG
d3	2	INT_TYPE, REAL_TYPE, BOOL_TYPE
d4	2	ID
d5	3	END_OF_INSTR
d6	3	ASSIGN
d7	4	LESS
d8	4	GREATER
d9	4	EQUALS
d10	4	FOR_WHILE_INSTR
d11	5	ID
d12	5	INT_VAL, REAL_VAL, BOOL_VAL
d13	6	SUM
d14	6	SUBSTRACT
d15	6	PRODUCT
d16	6	DIVIDE
d17	6	LESS, GREATER, EQUALS, ID, INT_VAL, REAL_VAL BOOL_VAL, END_OF_INSTR
d18	7	INT_VAL
d19	7	REAL_VAL
d20	7	BOOL_VAL
d21	8	INT_TYPE
d22	8	REAL_TYPE
d23	8	BOOL_TYPE

Conjuntos directores

5. Tabla de símbolos

5.1. Introducción

La tabla de símbolos almacena la información que en cada momento se necesita sobre las variables del programa y otra información, como el nombre, el tipo, el tamaño... También sirve para guardar información referente a cualquier clase, enumerado, identificador, etc. creado por el usuario.
La gestión de la tabla de símbolos es muy importante, ya que consume gran parte del tiempo de compilación; de ahí que su eficiencia sea crítica.
Respecto a las variables de usuario, la información que puede ser almacenada en la tabla de símbolos se divide en:

Nombre Se puede hacer con o sin límite. Si se hace con límite, emplearemos una longitud fija para cada variable, lo cual aumenta la velocidad de creación, pero limita la longitud en unos casos, y desperdicia espacio en la mayoría. Otro método es habilitar la memoria que necesitemos en cada caso para guardar el nombre.

Tipo Tipo de dato de la variable.

Dirección de memoria Esta dirección es necesaria, porque las instrucciones que referencian a una variable deben saber donde encontrar el valor de esa variable en tiempo de ejecución, también cuando se trata de variables globales.

Número de dimensiones de una variable array, o el de parámetros de una función o procedimiento Puede ser útil para el chequeo semántico.

La tabla de símbolos también resulta útil para almacenar información relativa a las constantes, palabras reservadas y funciones de librería. Conforme van apareciendo nuevas declaraciones de identificadores, el analizador léxico, o el analizador sintáctico según la estrategia que sigamos, insertará nuevas entradas en la tabla de símbolos, evitando siempre la existencia de entradas repetidas.

5.2. Operaciones de la tabla de símbolos

La tabla de símbolos consta de una estructura llamada símbolo. Las operaciones que puede realizar son:

Crear Crea una tabla vacía.

Insertar Añade un nodo a la cabeza de la tabla.

Buscar Busca el nodo que contiene el nombre pasado por parámetro.

Imprimir Devuelve una lista con los valores que tiene los identificadores de usuario, recorriendo la tabla de símbolos.

5.3. Ejemplo de uso de la tabla de símbolos

Suponemos que queremos hacer las siguientes operaciones:
a = 7 * 3
b = 3 * a
a = a + b

En la segunda instrucción necesitamos saber cuanto vale 'a'; es decir, el valor de 'a' debe estar guardado en algún sitio. Para ello utilizaremos una lista de pares nombre, valor; de forma que, cuando nos encontremos con la instrucción a = 7 * 3, miremos en la tabla y, si no está 'a' en la tabla, creamos un nodo para introducirla.

Nombre	**Valor**
a	21

A continuación nos encontramos con b = 3 * a. Buscamos 'a' en la tabla de símbolos y vemos que el valor de 'a' es 21. Ahora buscamos 'b' en la tabla de símbolos y como no está lo creamos.

Nombre	**Valor**
a	21
b	63

Ejecutamos ahora la instrucción a = a + b. Tendríamos a 'a' y 'b' en la tabla de símbolos con lo cual solo tendríamos que modificar el valor de 'a'.

Nombre	**Valor**
a	84
b	63

6. Generación de código