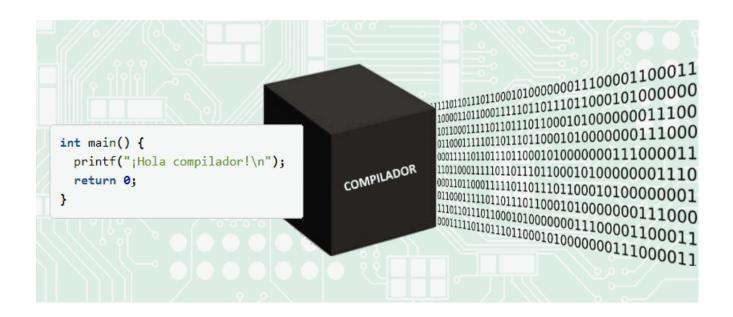
Diseño de Compiladores 1

Grupo 10 - Tp 1 y 2



Integrantes

Medioli Nicolas - mediolifnicolas@gmail.com

Arias Facundo - facundoarias1231@gmail.com

Docente asignado

Jose Fernandez Leon

Temas asignados

6 8 12 14 18 20 21 24 28 29 32 35

índice

Introducción	3
Especificaciones asignadas	
Parte 1 - Analizador Léxico	7
Estructura general del analizador léxico	8
Implementación general - LexycalAnalizer	9
Diagrama de transición de estados	9
Acciones semánticas	11
Errores léxicos considerados	13
Parte 2 - Analizador Sintáctico	14
Especificaciones generales	14
implementaciones generales - Yacc	15
Gramática	15
Lista de no terminales	16
Errores sintácticos considerados	18
Casos de prueba	19
Casos de prueba-Léxico	19
Casos de prueba-Sintáctico	27
Correcciones realizadas - 2da entrega	37
Conclusión	38

Introducción

En este informe se detalla el desarrollo y la implementación de un compilador y las dos primeras partes que lo conforman: Análisis Léxico y Análisis Sintáctico. Las especificaciones del mismo, fueron dadas por la cátedra.

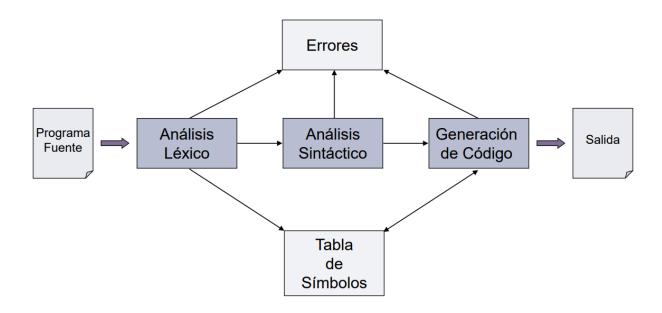


Figura 1: "Estructura general del compilador"

El ejecutable del compilador se ejecuta abriendo una consola dentro de la carpeta "entrega grupo 10" y escribiendo el siguiente comando de ejemplo:

java -jar build\compilador.jar tests\lexico\case1.fn

Especificaciones asignadas

- **6.** Enteros largos (32 bits): Constantes enteras con valores entre –2 31 y 231 1. Estas constantes llevarán el sufijo "_l".
- **8.** Punto flotante de 64 bits: Números reales con signo y parte exponencial. La parte exponencial puede estar ausente. Si está presente, el exponente comienza con la letra "D" (mayúscula o minúscula) y el signo del exponente es obligatorio. Puede estar ausente la parte entera o la parte decimal, pero no ambas. El "." es obligatorio.

Ejemplos válidos: 1. .6 -1.2 3.d-5 2.D+34 2.5D-1 13. 0. 1.2d+10

Considerar el rango 2.2250738585072014D-308 < x < 1.7976931348623157D+308 -1.7976931348623157D+308 < x < -2.2250738585072014D-308

Se debe incorporar a la lista de palabras reservadas la palabra DOUBLE.

- **12.** En las sentencias de asignación, puede utilizarse el operador '-=' en lugar del operador de asignación '='. Por ejemplo: a -= 10 i, z -= a * b,
- **14.** DO UNTIL DO UNTIL (), tendrá la misma definición que la condición de las sentencias de selección. podrá contener una sentencia, o un grupo de sentencias ejecutables delimitadas por llaves.
- **18.** Herencia por Composición Uso con nombre

Incorporar, como sentencia declarativa, la declaración de clases con la estructura que se muestra en los siguientes ejemplos:

Incorporar, como sentencia declarativa, la declaración de objetos de una clase determinada:

```
ca a1; a2,
cb b1; b2; b3,
```

Incorporar, dentro de las sentencias ejecutables, la posibilidad de utilizar referencias a métodos y atributos de objetos indicando explícitamente cuando el atributo o método corresponden a la clase heredada por composición.

Ejemplos:

```
a1.a = 3_i,
b1.ca.a = 3_i,
b1.a = 2.3,
b1.b = 1.2,
b1.ca.c = 1_i,
b1.ca.m(),
b1.n(),
```

(todos los chequeos semánticos para estas referencias, se efectuarán en la etapa 3 del TP)

20. Declaración de métodos distribuida La implementación de los métodos de una clase puede efectuarse mediante una cláusula IMPL como se indica en el siguiente ejemplo

// LÉXICO: Incorporar a los símbolos detectados, el símbolo ':'.

21. Forward declaration

- **24.** Sobreescritura de atributos La semántica de este tema se explicará en la etapa 3 del Trabajo Práctico
- 28. La semántica de estos temas se explicará en la etapa 3 del Trabajo Práctico
- **29.** Conversiones Explícitas: Se debe incorporar en todo lugar donde pueda aparecer una expresión, la posibilidad de utilizar la siguiente sintaxis:

TOF () // para grupos que tienen asignado el tema 7

TOD() // para grupos que tienen asignado el tema 8

//LÉXICO: Incorporar a la lista de palabras reservadas, la palabra TOF o TOD según corresponda.

- 32. Comentarios de 1 línea: Comentarios que comiencen con "**" y terminen con el fin de línea.
- **35.** Cadenas multilínea: Cadenas de caracteres que comiencen y terminen con " % " . Estas cadenas pueden ocupar más de una línea. (En la Tabla de símbolos se guardará la cadena sin los saltos de línea).

Ejemplo: %;Hola

mundo! %

Parte 1 - Analizador Léxico

Para desarrollar el analizador léxico tuvimos en cuenta el objetivo y las especificaciones que proporciona la cátedra:

Objetivo

Desarrollar un Analizador Léxico que reconozca los siguientes tokens:

- Identificadores cuyos nombres pueden tener hasta 20 caracteres de longitud. El primer puede ser una letra o '_', y
 el resto pueden ser letras, dígitos y "_". Los identificadores con longitud mayor serán truncados y esto se
 informará como Warning. Las letras utilizadas en los nombres de identificador sólo pueden ser minúsculas.
- Constantes correspondientes al tema particular asignado a cada grupo.
 Nota: Para aquellos tipos de datos que pueden llevar signo, la distinción del

Nota: Para aquellos tipos de datos que pueden llevar signo, la distinción del uso del símbolo '-' como operador aritmético o signo de una constante, se postergará hasta el trabajo práctico Nro. 2.

- Operadores aritméticos: "+", "-", "*", "/" agregando lo que corresponda al tema particular.
- Operador de asignación: "="
- Comparadores: ">=", "<=", ">", "<", "==", "!!"
- "{", "}", "(", ")",","," y ";"
- Cadenas de caracteres correspondientes al tema particular de cada grupo.
- Palabras reservadas (en mayúsculas):

```
IF, ELSE, END_IF, PRINT, CLASS, VOID
```

y demás símbolos / tokens indicados en los temas particulares asignados a cada grupo.

El Analizador Léxico debe eliminar de la entrada (reconocer, pero no informar como tokens al Analizador Sintáctico), los siguientes elementos.

- · Comentarios correspondientes al tema particular de cada grupo.
- · Caracteres en blanco, tabulaciones y saltos de línea, que pueden aparecer en cualquier lugar de una sentencia.

Analizador Léxico. Especificaciones

- a) El Analizador Léxico deberá leer un código fuente, identificando e informando:
 - Tokens detectados en el código fuente. Por ejemplo:

```
Palabra reservada IF
(
| Identificador var_x +
Constante entera 25
Palabra reservada ELSE
```

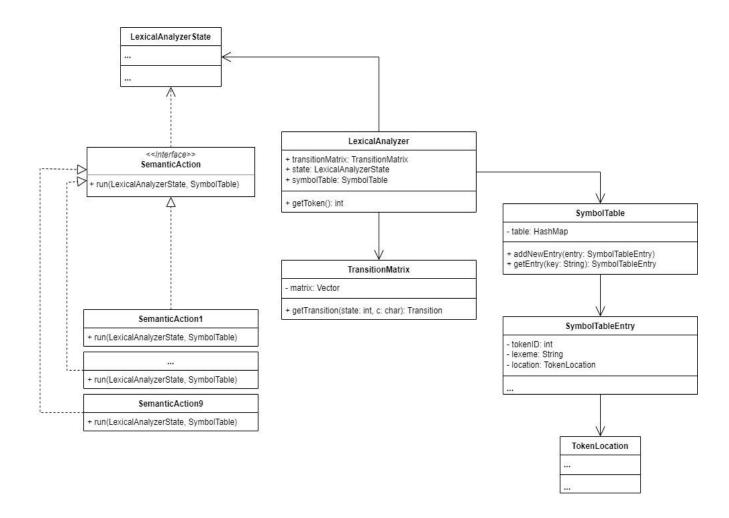
- Errores léxicos detectados en el código fuente, indicando: nro. de línea y descripción del error. Por ejemplo:
 Línea 24: Constante entera fuera del rango permitido
- Contenidos de la Tabla de Símbolos.

Se sugiere la implementación de un consumidor de tokens que invoque al Analizador Léxico solicitándole tokens. En el trabajo práctico 2, esta funcionalidad estará a cargo del Analizador Sintáctico.

- b) El código fuente debe ser leído desde un archivo, cuyo nombre debe poder ser elegido por el usuario del compilador. Se espera que el compilador pueda ejecutarse desde línea de comandos.
- c) La numeración de las líneas de código debe comenzar en 1. De este modo, la información de cada error coincidirá con el número de línea en el archivo del código fuente.
- d) Para la programación se podrá elegir el lenguaje. Para esta elección, tener en cuenta que el analizador léxico se integrará luego a un Parser (Analizador Sintáctico) generado utilizando una herramienta tipo Yacc. Por lo tanto, es necesario asegurarse la disponibilidad de dicha herramienta para el lenguaje elegido.
- e) El Analizador Léxico deberá implementarse mediante una matriz de transición de estados y una matriz de acciones semánticas, de modo que cada cambio de estado y acción semántica asociada, sólo dependa del estado actual y el carácter leído.
- f) Implementar una Tabla de Símbolos donde se almacenarán identificadores, constantes, y cadenas de caracteres. Es requisito para la aprobación del trabajo, que la tabla sea implementada con una estructura dinámica.
- g) La aplicación deberá mostrar, además de tokens y errores léxicos, los contenidos de La Tabla de Símbolos.

Decidimos utilizar el lenguaje Java porque es el lenguaje que más conocemos en común y se complementa fácilmente con Yacc, teniendo mucha documentación a disposición.

Estructura general del analizador léxico



Implementación general - LexycalAnalizer

En primera instancia creamos una clase "Compilador" que llama a otra clase "LexicalAnalyzer", esta última tiene como atributos la matriz de transición, la tabla de símbolos, un arreglo con acciones semánticas y un estado actual, entre otros. Luego se inicia la clase (asignando sus respectivos valores cada atributo). Esta, tiene como método principal "getToken", el mismo, mientras esté leyendo el token, hará una serie de funciones, entre ellas, pasar al siguiente estado, ejecutar acciones semánticas, etc.

Otra clase que es importante mencionar es "LexicanAnalyzerState", esta es una instancia del analizador léxico que tiene como atributos el índice del archivo que se está leyendo, estado y lexema actual, una variable booleana que te indica si estas leyendo el token, el token a devolver, etc. En resumen, funciona como un registro que contiene los datos actuales del token que se está leyendo.

La tabla de símbolos es implementada con un HashMap, donde la clave es generada de manera aleatoria a partir de un UUID, y cada "SymbolTableEntry" tiene información de:

- El tokenID
- El lexema
- La ubicación

Por último, esta clave es utilizada para guardar en el yylval la entrada a la información del token reconocido.

Diagrama de transición de estados

En primer lugar, debíamos realizar el diagrama de transición de estados. Utilizando <u>draw.io</u> creamos todos los posibles estados y sus respectivas transiciones hasta llegar al nodo final (por una cuestión de mejor legibilidad, hay dos nodos finales).

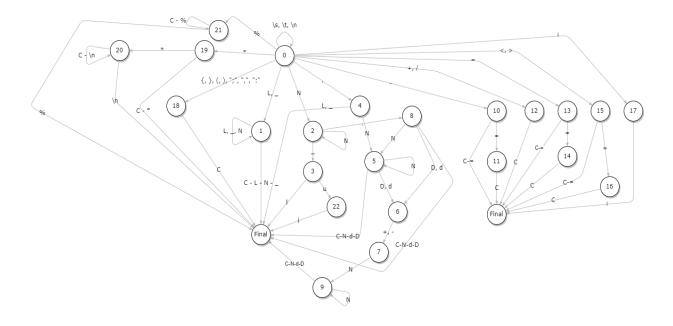


figura 2: "Diagrama de transición de estados"

Para implementar este diagrama en código, se podría haber utilizado una matriz de transición de estados como la que aparece en la teoría, pero tomamos la decisión de volver a utilizar un HashMap como aparece a continuación:

```
public class TransitionMatrix {
    private static Vector<HashMap<Character, Transition>> matrix;
```

Luego, cargamos el HashMap con los datos que aparecen en el diagrama (*figura 2*). Decidimos utilizar esta estructura porque nos pareció eficiente y por las mismas razones que se dieron anteriormente en la tabla de símbolos.

El componente Transition que tiene el HashMap tiene dos atributos, uno es el estado nuevo al que se accede cuando recibe un carácter y el otro es una lista de acciones semánticas que se ejecutan.

```
public class Transition {
    private Integer newState;
    private List<Integer> semanticActionList;
```

figura 3: "Atributos de la clase Transition"

Por ejemplo:

si el analizador detecta el carácter '_ ' se pasará al estado 1 (como se marca en el diagrama de estados) ejecutándose la acción semántica 2:

```
HashMap<Character, Transition> s0 = new HashMap<>();
s0.put(key:'_', new Transition(newState:1, List.of(e1:2)));
```

figura 4: inicialización de HashMap y ejemplo de inserción de dupla

Las acciones semánticas están cargadas en un arreglo en la clase "LexicalAnalyzer.java", entonces cuando se quiera ejecutar alguna acción, simplemente hay que acceder al arreglo en el índice que indica la transición y ejecutarla.

Acciones semánticas

En total generamos 9 acciones semánticas, estas, implementan un método que implementan la interfaz SemanticAction (método run), que recibe como parámetro el estado actual del léxico (anteriormente mencionado) y la tabla de símbolos.

Acción semántica 1: Se encarga de decrementar el índice de lectura del archivo.

- Acción semántica 2: Se encarga de resetear el lexema actual sumándole el último carácter leído.
- Acción semántica 3: A diferencia de la acción anterior, esta, simplemente le suma el último carácter leído al lexema actual. Pero si el carácter leído es un salto de línea, se omite.
- Acción semántica 4: Si el lexema tiene más de 20 caracteres, es truncado avisando al usuario mediante un warning. Luego lo busca en la tabla de símbolos predefinidos, y si está, devuelve el token junto a su ubicación. En caso de que no se encuentre, se agrega a la tabla de símbolos dinámica (de ahora en adelante, simplemente tabla de símbolos) el token 'ID', su ubicación y el lexema. Finalmente se retorna el token y su referencia en la tabla de símbolos
- **Acción semántica 5:** Verifica el rango de la constante y la da de alta en la tabla de símbolos, luego devuelve el token y su referencia en la tabla.
- Acción semántica 6: Devuelve un token predefinido
- Acción semántica 7: Resetea el lexema actual para comenzar con una nueva lectura.
- **Acción semántica 8:** Agrega el string (presente en el lexema actual) a la tabla de símbolos y retorna el token 'CTE_STRING' junto a su referencia en la tabla.

• *Acción semántica 9:* Incrementa el índice de línea que se está leyendo del archivo. Para incluir la información de ubicación de los tokens en la tabla de símbolos.

Errores léxicos considerados

- Enteros largos de 32 bits que no tengan como sufijo "_l"
- Punto flotante:
 - o no contener "."
 - o si se declara la parte exponencial, no contener "d" o "D"
- Comentarios:
 - o no comenzar con "**"
 - o no finalizar con un salto de línea
- Las cadenas de caracteres que no empiezan y terminan con "%"
- Chequeo de rangos para las constantes.
- Identificadores cuyos nombres tengan más de 20 caracteres (Warning).

Parte 2 - Analizador Sintáctico

Al igual que el analizador léxico, tuvimos en cuenta las especificaciones generales dadas por la cátedra para realizar la parte sintáctica:

OBJETIVO

Construir un Parser (Analizador Sintáctico) que invoque al Analizador Léxico creado en el Trabajo Práctico Nº 1, y que reconozca un lenguaje con las siguientes características:

SINTAXIS GENERAL:

Programa:

- Programa constituido por un conjunto de sentencias, que pueden ser declarativas o ejecutables.
 Las sentencias declarativas pueden aparecer en cualquier lugar del código fuente, exceptuando los bloques de las sentencias de control.
 - Los elementos declarados sólo serán visibles a partir de su declaración (esto será chequeado en etapas posteriores).
- El programa estará delimitado por llaves '{ y }'.
- Cada sentencia debe terminar con coma ",".

Sentencias declarativas:

 Sentencias de declaración de datos para los tipos de datos correspondientes a cada grupo según la consigna del Trabajo Práctico 1, con la siguiente sintaxis:

```
<tipo> <lista_de_variables>,
```

Donde <tipo> puede ser (Según tipos correspondientes a cada grupo): SHORT, INT, LONG, USHORT, UINT, ULONG, FLOAT, DOUBLE

Las variables de la lista se separan con punto y coma (";")

Incluir declaración de funciones VOID, con la siguiente sintaxis:

```
VOID ID (<parametro>)
{
<cuerpo_de_la_funcion>
}
```

Donde:

<parametro> será un identificador precedido por un tipo.:

```
<tipo> ID
```

Se permite hasta un parámetro, y puede no haber parámetros. Este chequeo debe efectuarse durante el Análisis Sintáctico

 <uerpo_de_la_funcion> es un conjunto de sentencias declarativas (incluyendo declaración de otras funciones) y/o ejecutables, incluyendo sentencias de retorno con la siguiente estructura:

RETURN,

Eiemplos válidos:

Sentencias ejecutables:

Asignaciones donde el lado izquierdo puede ser un identificador, y el lado derecho una expresión aritmética.
 Los operandos de las expresiones aritméticas pueden ser variables, constantes, u otras expresiones aritméticas

No se deben permitir anidamientos de expresiones con paréntesis.

· invocación a una función, con el siguiente formato:

 Cláusula de selección (IF). Cada rama de la selección será un bloque de sentencias. La estructura de la selección será, entonces:

```
IF (<condicion>) <bloque_de_sent_ejecutables> ELSE <bloque_de_sent_ejecutables> END_IF,
```

El bloque para el ELSE puede estar ausente.

La condición será una comparación entre expresiones aritméticas, variables o constantes, y debe escribirse entre "(" ")".

El bloque de sentencias ejecutables puede estar constituido por una sola sentencia, o un conjunto de sentencias ejecutables delimitadas por llaves.

Sentencia de salida de mensajes por pantalla. El formato será

```
PRINT<cadena>,

Ejemplos:

PRINT#Hola mundo#, //Tema 34

PRINT %Hola //Tema 35
```

Implementación General - Yacc

Para implementar el analizador sintáctico, utilizamos Yacc para analizar estructuralmente una entrada, esta herramienta requiere 3 especificaciones:

- Una gramática (Conjunto de reglas)
- Código a ser invocado cuando una regla es reconocida
- Un analizador léxico que le provea los tokens

Gramática

Tuvimos que realizar una gramática que reconozca las reglas, esta está contenida en un archivo llamado "compilador.y".

En primer lugar declaramos como tokens las palabras reservadas que actúan como estados finales (además de los tokens con relación 1:1 con ascii) y luego definimos la gramática.

```
%token IF ELSE END_IF PRINT CLASS VOID ID

LONG UINT DOUBLE STRING

CTE_LONG CTE_UINT CTE_DOUBLE CTE_STRING

CMP_GE CMP_LE CMP_EQUAL CMP_NOT_EQUAL

SUB_ASIGN

DO UNTIL IMPL FOR RETURN TOD
```

"Figura 5: Definición de palabras reservadas"

Para poder mostrar por pantalla el conjunto de reglas detectados en el código, utilizamos una cola de prioridad en la que se almacenan las diferentes estructuras sintácticas encontradas (SyntacticStructureResult) en base a su ubicación en el archivo.

Esta implementación nos permite imprimir de forma ordenada el conjunto de reglas sintácticas que reconoce el parser.

Lista de no terminales

- programa: encierra todo el programa entre {}.
- comparador: operadores de comparación.
- condicion: comparación de expresiones.
- tipo: incluye palabras reservadas de los tipos de datos.
- lista identificadores: lista recursiva de identificadores.
- sentencia declarativa: incluye las sentencias de:
 - declaración de variable
 - definición de función y clases
 - o implementación de métodos distribuidos.
- sentencia ejecutable:
 - asignación a variables/atributos
 - invocación a función/método
 - sentencia if, do until, PRINT y RETURN
- lista_sentencias: lista de sentencias de cualquier tipo.
- lista_sentencias_ejecutables: sólo sentencias ejecutables.
- invocacion funcion: invocación a función con o sin parámetro.
- op_asignacion_aumentada: asignación con = o -=.
- sentencia_if: sentencia condicional.
- constante: constante de cualquier tipo de dato.

- expr: expresión con casting.
- basic_expr: expresión sin anidamiento.
- term: término de una expresión.
- factor: factor de una expresión.
- parametro_formal: sintaxis de parámetro en declaración de funciones.
- parametro_real: sintaxis de parámetro en invocación a funciones.
- definicion_funcion: definición de función.
- procedimiento: definición de procedimiento (función o método)
- do until: sentencia de bucle condicional
- metodo: definición de método.
- acceso_atributo: acceso a atributos mediante operador de acceso.
- definicion_clase: definición de clase.
- cuerpo_clase: cuerpo de clase según la especificación.
- clase_lista_atributos: lista de atributos dentro de clase.
- clase_lista_metodos: lista de métodos dentro de clase.
- clase_lista_composicion: lista de clases para composición.
- implementacion: implementación de métodos de forma distribuida.

Errores sintácticos considerados

• Sentencias que no terminen con coma ','

- Programa que no esté delimitado por llaves '{' '}'
- Sentencias declarativas:
 - o Lista de variables declaradas que no estén separadas con un punto y coma ';'
 - Declaración de funciones:
 - Parámetro que no sea de la forma <tipo> ID
 - Cuerpo de la función que no esté delimitado con llaves '{' '}'
 - Funciones con mas de 1 parametro
- Sentencias ejecutables:
 - o Asignaciones donde el lado izquierdo no sea un identificador
 - Anidamiento de expresiones con paréntesis
 - Cláusula de selección IF:
 - Ramas de selección que no sean bloques de sentencias
 - Condición que no esté delimitada con paréntesis '(' ')'
 - Bloques de sentencias que no estén delimitadas por llaves '{' '}'
 - Símbolo de negativo en constantes tipo UINT

Casos de prueba

Para manejar y mostrar por pantalla los errores detectados, decidimos guardar como información adicional en la tabla de símbolos la línea donde se encuentran cada token, por lo tanto cuando se encuentre un error de cualquier tipo, se puede identificar de qué línea surge y mostrarla.

Luego de las incorporaciones nuevas agregadas en las partes correspondientes a la generación de código intermedio y código assembler, a la hora de probar la salida del compilador, se van a mostrar datos, variables que no se explican en este informe, código assembler, etc (ya que posteriormente son explicadas). Se espera que el lector haga énfasis en la correspondiente impresión de la tabla de símbolos y las estructuras sintácticas encontradas.

Casos de prueba - Léxico

 Case 1: para cada tipo de datos asignado, probamos si funciona el primer y último valor del rango. Esperamos que funcione bien y muestre por pantalla los siguiente para el siguiente código:

```
Constantes con el primer y ultimo valor dentro del rango (Para cada tipo de datos asignado).
** Se espera que funcione BIEN
   LONG var1,
   UINT var2,
   DOUBLE var3,
    ** tipo long
   var1 = -2147483648_1,
    ** tipo long
   var1 = 2147483647 1,
    ** tipo uint
   var2 = 0 ui,
    ** tipo uint
   var2 = 65535_ui,
    ** tipo double
   var3 = -1.7976931348623157d + 308.
    ** tipo double
    var3 = 1.7976931348623157d+308,
```

```
Parsing correcto
Lista de tokens leidos:
'('' 1D '=''-''-' CTE_DOUBLE ',' 1D '=' CTE_DOUBLE ',' 1D '=' CTE_UINT ',' 1D '=' CTE_UINT ',' 1D '=' '-' TOKEN NO COMOCIDO ',' 1)'

Setructuras sintacticae encontradas:
Lenea S, estructura: Asignacion a variable
Linea A, estructura: Asignacion a variable
Linea A, estructura: Asignacion a variable
Linea II, estructura: Asignacion a variable
Lexema: '-1.7976931348623157d+308'
Lexema: '1.7976931348623157d+308'
Lexema: '1.797693134623157d+308'
Lexema: '1.797693134623157d+308'
Lexema: '
```

 Case 2: Similar a la anterior, pero los valores asignados están fuera del rango, por lo tanto se espera que funcione mal, mostrando:

```
** Constantes con el primer y ultimo valor fuera del rango (Para cada tipo de datos asignado).
     ** Se espera que funcione MAL
 4
         LONG var1,
         UINT var2,
         DOUBLE var3,
 7
         ** tipo LONG
         var1 = -2147483649_1,
         ** tipo LONG
         var1 = 2147483648 1,
         ** tipo UINT
         var2= -1_ui,
         ** tipo UINT
         var2 = 65536_ui,
         ** tipo DOUBLE
         var3 = -1.7976931348623159d+308,
         ** tipo DOUBLE
         var3 = 1.7976931348623159d+308,
25
```

```
Section of the control of the contro
```

 Case 3: para los números de punto flotante, parte entera con y sin parte decimal, parte decimal con y sin parte entera, con y sin exponente y exponente positivo o negativo. Debe funcionar bien mostrando:

```
** Para numeros de punto flotante
** Se espera que funcione BIEN

| DOUBLE variable,
| variable = 2442.,
| variable = 2442.523,
| variable = .634,
| variable = 2442.45d+24,
| variable = 2442.45d-24,
| variable = .45d-24,
| variable = .45d-24,
| variable = .45d+24,
```

 Case 4: Probamos Identificadores de menos y más de 25 caracteres, se espera que funcione bien truncando los identificadores largos y mostrando:

```
** Identificadores de menos y mas de 25 caracteres.

** Se espera que funcione BIEN (con warning)

{
LONG identificador_corto,
LONG identificador_muy_largo,

identificador_corto = 25_l,
identificador_muy_largo = 23_l,
}
```

```
[Linea 5] Lexema truncado
identificador_muy_largo -> identificador_muy_la
Parsing correcto
Lista de tokens leidos:
'{' ID '=' CTE_DOUBLE ',' ID '=' CTE_DOUBLE ',' '}'
Estructuras sintacticas encontradas:
Linea 4, estructura: Asignacion a variable
Linea 5, estructura: Asignacion a variable
Tabla de simbolos:
Clave: 66381c | To
                              TokenID: 125
TokenID: 263
TokenID: 61
                                                             Ubicacion: Linea 6
 lave: 283c5b
lave: a20df9
lave: 58f12f
lave: 445c18
                                                             Ubicacion: Linea
Ubicacion: Linea
                                                                                                            Lexema: 'identificador_muy_la'
                              TokenID: 123
TokenID: 44
                                                             Ubicacion: Linea
                                                             Ubicacion: Linea
 lave: 58ea5e
lave: 582b5d
                              TokenID: 268
                                                             Ubicacion: Linea
                                                                                                            Lexema: '23_1'
                             TokenID: 268
TokenID: 268
TokenID: 61
                                                             Ubicacion: Linea 5
Ubicacion: Linea 4
                                                                                                            Lexema: 'identificador_corto'
Lexema: '25_1'
 lave: f08131
 clave: 100131 | TokenID: 203 | Obicacion: Linea 4 | Lexema: 'Tocheri
Clave: 53bbc5 | TokenID: 268 | Ubicacion: Linea 4 | Lexema: '25_1'
Clave: 0da5f9 | TokenID: 61 | Ubicacion: Linea 5
C:\Users\facun\0neDrive\Escritorio\entrega grupo 10\entrega grupo 10>
```

• Case 5: identificadores con letras, dígitos y guiones(_). Se espera que funcione bien:

```
** Identificadores con letras, digitos y _

** Se espera que funcione BIEN

{
    STRING id_con_guiones_y_num,
    UINT id_25,
    id_con_guiones_y_num = %test%,
    id_25 = 25_ui,

}
```

 Case 6: Similar al anterior, pero uso Identificadores con caracteres inválidos (que no sean letras, dígitos o guiones). Debe funcionar mal y mostrar:

```
** Intento de incluir en el nombre de u
** Se espera que funcione MAL

| Variab*e, |
| Variab*e = %hola%, |
|
```

 Case 7: Utilizamos palabras reservadas escritas en minusculas y mayusculas. Se espera que funcione bien:

```
** Palabras reservadas escritas en
** Se espera que funcione BIEN

{
    LONG do,
    STRING string,
    STRING return,

    do = 25_1,
    return = %hola%,
    string = %string%,

10    string = %string%,

11 }
```

• Case 8: Comentarios bien y mal escritos, por lo tanto se espera que funcione mal:

• Case 9: Cadenas bien y mal escritas, por lo tanto se espera que funcione mal:

```
PS C:\Users\facun\OneDrive\Escritorio\entrega grupo 10\entrega grupo 10> java -jar .\build\compilador.jar .\tests\lexico\case9.fn
[Lexico: Linea 6] Caracter no reconocido: '"'

Hubo errores en el parsing

Lista de tokens leidos:
'{' ID '=' CTE_STRING ',' ID '=' ID ID

Estructuras sintacticas encontradas:
Linea 4, estructura: Asignacion a variable

Tabla de simbolos:
Clave: 189905 | TokenID: 61 | Ubicacion: Linea 4
Clave: b06312 | TokenID: 123 | Ubicacion: Linea 3
Clave: b06312 | TokenID: 21 | Ubicacion: Linea 4 |
Clave: b06317 | TokenID: 21 | Ubicacion: Linea 4 |
Lexema: 'esta es una cadena
bien escrita'

Clave: cd96f6 | TokenID: 44 | Ubicacion: Linea 5
Clave: b2557c | TokenID: 263 | Ubicacion: Linea 6 |
Clave: g84c14 | TokenID: 263 | Ubicacion: Linea 6 |
Clave: 330b1f | TokenID: 263 | Ubicacion: Linea 6 |
Clave: 23cbcc | TokenID: 263 | Ubicacion: Linea 6 |
Exema: 'str_mal_escrita'
Clave: 23cbcc | TokenID: 261 | Ubicacion: Linea 6 |
Exema: 'str_mal_escrita'
Clave: 23cbcc | TokenID: 61 | Ubicacion: Linea 6 |
Exema: 'str_mal_escrita'
```

Casos de prueba - Sintáctico

 Case 1: Función VOID que imprime por pantalla un "hola". Se espera que funcione bien:

```
** Se espera que ande BIEN

VOID saludar()

PRINT %hola%,

saludar(),

saludar(),

}
```

• Case 2: Declaración de clase con sus respectivos atributos y métodos. Se espera que ande bien:

```
Passing correcto
Lista de tokens leidos:
'{ CLASS ID '{ STRING ID ', VOID ID '(' STRING ID ')' '{ ID '=' ID ', '}' ', ']' ', 'ID ID ', 'ID ', 'ID ', ']'
Estructuras sintactics encontradas:
Linea 3, estructura: Definicion de clase
Linea 7, estructura: Definicion de clase
Linea 8, estructura: Definicion de variable
Linea 1, estructura: Asignacion a variable
Linea 13, estructura: Asignacion a variable
Linea 13, estructura: Asignacion a variable
Linea 14, estructura: Impocacion a metodo
Linea 14, estructura: Invocacion a metodo
Linea 14, estructura: Asignacion a variable
Linea 15, estructura: Linea 16
Linea 17, estructura: Linea 17
Linea 18, estructura: Linea 18
Lexema: 'set_nombre'
Lave: 56367 | TokenID: 263 | Ubicacion: Linea 14
Lexema: 'set_nombre'
Lave: 67369 | TokenID: 263 | Ubicacion: Linea 14
Lexema: 'set_nombre'
Lave: 67369 | TokenID: 263 | Ubicacion: Linea 14
Lexema: 'set_nombre'
Lave: 67369 | TokenID: 263 | Ubicacion: Linea 14
Lexema: 'policion 10, est 10,
```

• Case 3: Cláusula de selección IF con las dos ramas (else). Debe andar bien, mostrando:

• Case 4: probamos la cláusula DO UNTIL. Debería funcionar bien:

```
PS C:\Users\facun\OneDrive\Escritorio\entrega grupo 10\entrega grupo 10\en
```

• Case 5: Herencia por composición e implementación de método a la clase hija. Se espera que funcione bien:

```
** Se espera que ande BIEN
   CLASS persona
        STRING nombre,
        LONG dni,
    },
    CLASS empleado
       LONG id_empleado,
        persona p1, ** Herencia por composicion
    },
    IMPL FOR empleado:
        VOID enviar_encomienda()
            PRINT %se envia encomienda%,
        },
    },
   empleado emp1,
    emp1.enviar_encomienda(),
```

```
PS C.\VENERS\factor\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\texture\te
```

• Case 6: Función dentro de otra función, se espera que funcione bien:

```
PS C:\Users\facun\OneDrive\Escritorio\entrega grupo 10\entrega grupo 10> java -jar .\build\compilador.jar .\tests\sintactico\case6.fn

Parsing correcto

Lista de tokens leidos:

('' VOTI D' '') '' (VOID ID '(' LONG ID ')' '{ PRINT CTE_STRING ', ' }' ', ' ID '(' CTE_DOUBLE ')' ', ' ', ' ID '(' ')' ', ' '}'

Estructuras sintacticas encontradas;
Linea 3, estructura: Definicion de funcion
Linea 5, estructura: Definicion de funcion
Linea 7, estructura: Definicion de funcion
Linea 7, estructura: Invocacion a funcion
Linea 10, estructura: Invocacion a funcion
Linea 10, estructura: Invocacion a funcion
Linea 12, estructura: Invocacion a funcion
Linea 12, estructura: Invocacion a funcion
Linea 12, estructura: Invocacion a funcion
Tabla de simbolos:
Clave: 647baa | TokenID: 240 | Ubicacion: Linea 9
Lexema: 647baa | TokenID: 240 | Ubicacion: Linea 9
Lexema: 93895a | TokenID: 241 | Ubicacion: Linea 9
Lexema: 93895a | TokenID: 252 | Ubicacion: Linea 9
Lexema: 948880 | TokenID: 253 | Ubicacion: Linea 13
Clave: 948880 | TokenID: 254 | Ubicacion: Linea 13
Clave: 948880 | TokenID: 271 | Ubicacion: Linea 12
Clave: 948880 | TokenID: 271 | Ubicacion: Linea 13
Clave: 948880 | TokenID: 271 | Ubicacion: Linea 13
Clave: 948880 | TokenID: 271 | Ubicacion: Linea 13
Clave: 948880 | TokenID: 271 | Ubicacion: Linea 13
Clave: 948880 | TokenID: 44 | Ubicacion: Linea 13
Clave: 948880 | TokenID: 45 | Ubicacion: Linea 16
Clave: 95897 | TokenID: 123 | Ubicacion: Linea 16
Clave: 95897 | TokenID: 123 | Ubicacion: Linea 16
Clave: 95897 | TokenID: 124 | Ubicacion: Linea 16
Clave: 19560 | TokenID: 123 | Ubicacion: Linea 16
Clave: 19560 | TokenID: 123 | Ubicacion: Linea 16
Clave: 19560 | TokenID: 123 | Ubicacion: Linea 16
Clave: 19560 | TokenID: 123 | Ubicacion: Linea 16
Clave: 19560 | TokenID: 123 | Ubicacion: Linea 16
Clave: 19560 | TokenID: 124 | Ubicacion: Linea 16
Clave: 19560 | TokenID: 125 | Ubicacion: Linea 17
Clave: 19560 | TokenID: 124 | Ubicacion: Linea 16
Clave: 19560 | TokenID: 125 | Ubicacion: Linea 17
Clave: 19560 | TokenID: 125 | Ubica
```

 Case 7: IF con una condición inválida ya que el número 53 no es una constante válida por el lenguaje, por lo tanto se espera que funcione mal mostrando:

```
1  ** Se espera que no funcione
2  ** (53 no es una constante valida)
3  {
4     LONG a,
5     a = 100_l,
6
7     IF (a < 53)
8     {
9         PRINT %error%,
10     }
11     ELSE
12     {
13         PRINT %error%,
14     }
15     END_IF,
16 }</pre>
```

```
PS C:\Users\facun\OneDrive\Escritorio\entrega grupo 10\entrega grupo 10> java -jar .\build\compilador.jar .\tests\sintactico\case7.fn
[Extico: Linea 7] Caracter no reconocido: ')'
[Sintactico: Linea 7] Error en IF

Hubo errores en el parsing
Lista de tokens leidos:
''( LONG ID ', ID '= CTE_DOUBLE ', IF '(' ID '<' '{' PRINT CTE_STRING ', '}' ELSE '{' PRINT CTE_STRING ', '}' ESTRUCTURA SINTACTICAS ENCORTED A CONTROL DE CARACTERIO A CONT
```

• Case 8: Pasamos dos datos como parametro a un método, pero esto no es válido, ya sea para un método o una función, por lo tanto debería andar mal y mostrar:

```
| Scale | Company | Compan
```

Correcciones realizadas - 2da entrega

1. "El 'compilador.jar' no funciona en cualquier caso que se pruebe, ya sea con un código que tiene o no errores"

Ejecutable arreglado.

 "No hay indicación en el informe de cómo ejecutar el compilador '.jar', como lo generaron en cuanto a que versión de Java usar, etc"
 Indicación puesta en la introducción (Pag 3)

3. "No se explica en el informe, ni se dan ejemplos de cuales son los resultados que se reportan al ejecutar el compilador"

Casos de prueba (pág 20)

- 4. "No se informa como se reporta la línea donde se encuentra algún error y como se informa tal error. Tampoco se habla de la política de manejo de errores utilizada" No se explicó en el informe el manejo de errores, la salida por pantalla muestra los errores y en la línea donde se encuentra.
- 5. "No se da un ejemplo de como se listan las reglas, tabla de símbolos, etc por pantalla." Corregido
- 6. "Deben listar en la tabla de símbolos cuál es la entrada y que token representan más allá de TokenID. Validen nuevamente el diagrama de transición de estados para cubrir ese, y otros caso que detecten para arreglar."
 Corregido, y el diagrama de transición fue validado
- " El compilador no detecta palabras reservadas respecto a su identificación de mayúsculas y minúsculas; e.g. int VARIABLE,"
 No pudimos corregirlo
- 8. "No encuentro casos de testeo de todos los puntos asignados al grupo; e.g. Herencia por Composición."

Caso de pruebas que faltaban añadidos

 "Sería conveniente que trabajen con los archivos de testeo creados como ".txt" y no como ".fn", principalmente por comodidad al crearlos y editarlos."
 Corregido

Conclusión

Al principio de la cursada no sabíamos cómo se podría traducir un código fuente a un código ejecutable, pero a medida que avanzabamos en este trabajo empezamos a comprender cómo se implementa un compilador, profundizando en su estructura y sus partes, tuvimos que informarnos sobre ciertos temas, tomar decisiones de implementación y aprender a usar herramientas nuevas. Si bien nos quedan dos etapas más que debemos aprender y profundizar, ya hemos podido realizar las dos primeras partes del compilador, hemos aplicado técnicas aprendidas en la carrera, manejamos los errores de manera efectiva y documentamos el proceso.

Para finalizar, la experiencia de crear un software complejo como este, nos puso a prueba y hemos podido mejorar nuestras habilidades y capacidades adquiriendo conocimientos nuevos.