

Diseño de Compiladores I – Cursada 2024

Trabajo Práctico Nro. 1

La entrega se hará en forma conjunta con el Trabajo Práctico Nro. 2. Fecha de Entrega: A definir

Objetivo

Desarrollar un Analizador Léxico que reconozca los siguientes tokens:

- Identificadores cuyos nombres pueden tener hasta 15 caracteres de longitud. El primer carácter sólo puede ser una letra, 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 pueden ser minúsculas o mayúsculas.
- Constantes correspondientes al tema particular asignado a cada grupo.
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: “+”, “-”, “*”, “/”.
- Operador de asignación: “:=”
- Comparadores: “>=”, “<=”, “>”, “<”, “=”, “!=”
- “(”, “)”, “,”, “.” y “;”
- Cadenas de caracteres correspondientes al tema particular de cada grupo.
- Palabras reservadas (que pueden escribirse con mayúsculas o minúsculas):
IF, THEN, ELSE, BEGIN, END, END_IF, OUTF, TYPEDEF, FUN, RET
- 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

Implementación:

- a) 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.
- b) Se debe implementar una Tabla de Símbolos donde se almacenarán identificadores, constantes, y cadenas de caracteres. Esta tabla **debe** ser implementada con una estructura dinámica.

[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.](#)

Aplicación a desarrollar:

- a) No se requiere interfaz gráfica. Se recomienda la implementación de una aplicación que se ejecute desde consola.
- b) El código fuente a compilar **debe estar en un archivo** cuyo **nombre pueda ser elegido por el usuario**. La ruta del archivo donde está el código a compilar se debe pasar como parámetro al ejecutar la aplicación desde la consola.
- c) El Analizador Léxico deberá leer el código fuente, identificando tokens, y debe generar como salidas.

- 1) Tokens detectados en el código fuente. Por ejemplo:

```
Palabra reservada IF
(
Identificador var_x
+
Constante entera 25
Palabra reservada else
etc.
```

- 2) 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

- 3) Contenidos de la Tabla de Símbolos.

Estas salidas se pueden imprimir por consola y/o enviarlas a uno o varios archivos de salida que deben quedar en la misma carpeta del archivo compilado.

Lenguaje de desarrollo:

- a) 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.
- b) Si usan GitHub, debe ser privado.

Entrega

La entrega se pactará con el docente asignado al grupo. La asignación se publicará en el Aula Virtual de la materia.

El material entregado debe incluir:

- Ejecutable del compilador
- Código fuente completo del compilador
- Casos de prueba
- Informe

Informe:

Debe incluir:

- **NRO. DE GRUPO** e Integrantes. Incluir **DIRECCIONES DE CORREO** para contacto.
- Temas particulares asignados (esta información deberá repetirse en los informes de los trabajos prácticos subsiguientes).
- Introducción.
- Decisiones de diseño e implementación.
- Diagrama de transición de estados.
- Matriz de transición de estados.
- Descripción del mecanismo empleado para implementar la matriz de transición de estados y la matriz de acciones semánticas.
- Lista de acciones semánticas asociadas a las transiciones del autómata del Analizador Léxico, con una breve descripción de cada una.
- Errores léxicos considerados.

[Este informe deberá ser completado con las consignas indicadas en el Trabajo Práctico 2.](#)

Casos de Prueba

Se debe incluir, como mínimo, ejemplos que contemplen las siguientes alternativas:

(Cuando sea posible, agregar un comentario indicando el comportamiento esperado del compilador)

- Constantes con el primer y último valor dentro del rango (Para cada tipo de datos asignado).
- Constantes con el primer y último valor fuera del rango (Para cada tipo de datos asignado).
- Para números de punto flotante: parte entera con y sin parte decimal, parte decimal con y sin parte entera, con y sin exponente, con exponente positivo y negativo.
- Identificadores de menos y más de 15 caracteres.
- Identificadores con letras, dígitos y "_".
- Intento de incluir en el nombre de un identificador un carácter que no sea letra, dígito o "-".
- Palabras reservadas escritas en minúsculas y mayúsculas.
- Comentarios bien y mal escritos.
- Cadenas bien y mal escritas.

Temas particulares

Cada grupo de trabajo tendrá asignada una combinación de temas particulares.
La información de los temas asignados a cada grupo, estará disponible en el Aula Virtual de la materia.

- 1. **Enteros (16 bits):** Constantes enteras con valores entre -2^{15} y $2^{15} - 1$. Se debe incorporar a la lista de palabras reservadas la palabra **integer**.
- 2. **Enteros sin signo (16 bits):** Constantes con valores entre 0 y $2^{16} - 1$. Se debe incorporar a la lista de palabras reservadas la palabra **uinteger**.
- 3. **Enteros largos (32 bits):** Constantes enteras con valores entre -2^{31} y $2^{31} - 1$. Se debe incorporar a la lista de palabras reservadas la palabra **longint**.
- 4. **Enteros largos sin signo (32 bits):** Constantes enteras con valores entre 0 y $2^{32} - 1$. Se debe incorporar a la lista de palabras reservadas la palabra **ulongint**.
- 5. **Punto Flotante de 32 bits:** Números reales con signo y parte exponencial. La parte exponencial puede estar ausente. Si está presente, el exponente comienza con la letra “s” (minúscula) y el signo del exponente es opcional. Su ausencia implica signo positivo para el exponente.

La parte entera, la parte decimal, y el ‘.’ son obligatorios.

Ejemplos válidos: 1.0 0.6 -1.2 3.0s-5 2.0s+34 2.5s-1 15.0 0.0 1.2s10

Considerar el rango $1.17549435s-38 < x < 3.40282347s+38 \cup$
 $-3.40282347s+38 < x < -1.17549435s-38 \cup 0.0$

Se debe incorporar a la lista de palabras reservadas la palabra **single**.

- 6. **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” (minúscula) y el signo del exponente es opcional. Su ausencia implica signo positivo para el exponente.

La parte entera, la parte decimal, y el ‘.’ son obligatorios.

Ejemplos válidos: 1.0 0.6 -1.2 3.0d-5 2.0d+34 2.5d-1 13.0 0.0 1.2d10

Considerar el rango $2.2250738585072014d-308 < x < 1.7976931348623157d+308 \cup$
 $-1.7976931348623157d+308 < x < -2.2250738585072014d-308 \cup 0.0$

Se debe incorporar a la lista de palabras reservadas la palabra **double**.

- 7. **Constantes octales**, que se escribirán comenzando con un **0**, y serán una secuencia de dígitos que pueden ir de 0 a 7. El rango para las constantes octales será el mismo que el del tipo entero asignado al grupo:
 - a. Para integer: -010000 a 077777
 - b. Para uinteger: 00 a 0177777
 - c. Para longint: -020000000000 a 017777777777
 - d. Para ulongint: 00 a 037777777777
- 8. **Constantes hexadecimales**, que se escribirán comenzando con **0x**, y serán una secuencia de dígitos que pueden ir de 0 a 9 y letras de la A a la F. El rango para las constantes hexadeimales será el mismo que el del tipo entero asignado al grupo:
 - a. Para **integer**: -0x8000 a 0x7FFF
 - b. Para **uinteger**: 0x0 a 0xFFFF
 - c. Para **longint**: -80000000 a 0x7FFFFFFF
 - d. Para **ulongint**: 0x0 a 0xFFFFFFFF

- 9. **Cadenas de 1 línea:** Cadenas de caracteres delimitadas por llaves (estas cadenas no pueden ocupar más de una línea).

Ejemplo: { ¡Hola mundo ! }

- 10. **Cadenas multilínea:** Cadenas de caracteres delimitadas por corchetes. 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!]

- 11. Incorporar a la lista de palabras reservadas la palabra **TYPDEF**.
- 12. La primera letra de un identificador podrá definir el tipo de un identificador:
 - a. Para **integer**, los identificadores que comiencen con **i**, **j** o **k** serán de tipo entero
 - b. Para **longint**, los identificadores que comiencen con **l**, **m** o **n** serán de tipo entero largo
 - c. Para **uinteger**, los identificadores que comiencen con **u**, **v** o **w** serán de tipo entero sin signo
 - d. Para **ulongint**, los identificadores que comiencen con **x**, y o **z** serán de tipo entero largo sin signo
 - e. Para **single**, los identificadores que comiencen con **s** serán de tipo single.
 - f. Para **double**, los identificadores que comiencen con **d** serán de tipo double.
- 13. Incorporar a la lista de palabras reservadas, la palabra **WHILE**.
- 14. Incorporar a la lista de palabras reservadas, las palabras **REPEAT** y **UNTIL**.
- 15. Incorporar a la lista de palabras reservadas, las palabras **REPEAT** y **WHILE**.

- 16. Incorporar a la lista de palabras reservadas, las palabras **FOR**, **UP** y **DOWN**.
- 17. A definir en Trabajos Prácticos 2/3.
- 18. A definir en Trabajos Prácticos 2/3.
- 19. A definir en Trabajos Prácticos 2/3.
- 20. Incorporar a la lista de palabras reservadas la palabra **STRUCT**.
- 21. Incorporar a la lista de palabras reservadas la palabra **PAIR**.
- 22. Incorporar a la lista de palabras reservadas la palabra **TRIPLE**.
- 23. Incorporar a la lista de palabras reservadas la palabra **GOTO**.
- 24. A definir en Trabajos Prácticos 2/3.
- 25. A definir en Trabajos Prácticos 2/3.
- 26. A definir en Trabajos Prácticos 2/3.
- 27. A definir en Trabajos Prácticos 2/3.
- 28. **Comentarios de 1 línea:** Comentarios que comiencen con “##” y terminen con el fin de línea.
- 29. **Comentarios multilínea:** Comentarios que comiencen con “/#” y terminen con “#” (estos comentarios pueden ocupar más de una línea).