



UNIVERSIDAD DE
GUADALAJARA
Red Universitaria e Institución Benemérita de Jalisco



Alumno:

Romo Rodríguez José Alberto (216853747)

Materia:

I7028 | Seminario de Solucion de Problemas de Traductores de Lenguajes II

NRC:

103841

Sección:

Do2

Maestro:

Lopez Franco Michel Emanuel

Horario:

Lunes y Miercoles | 13:00 - 14:55

Tarea:

¿Que es un analizador léxico?

Fecha de Entrega:

23 de Enero del 2022

Analizador Léxico.

Un **analizador léxico** es la primera fase de un compilador, consistente en un programa que recibe como entrada el código fuente de otro programa, normalmente una secuencia de caracteres específicos, y produce una salida compuesta de *tokens* o símbolos que son componentes léxicos previamente establecidos en el compilador.

Estos *tokens* sirven para una posterior etapa del proceso de traducción, siendo la entrada para el analizador sintáctico, es decir realizar la tarea de parsing.

La especificación de un lenguaje de programación a menudo incluye un conjunto de reglas que definen el léxico. Estas reglas consisten comúnmente en expresiones regulares que indican el conjunto de posibles secuencias de caracteres que definen un *token* o lexema. En algunos lenguajes de programación es necesario establecer patrones para caracteres especiales (como el espacio en blanco) que la gramática pueda reconocer sin que constituya un *token* en sí.

Esta etapa está basada usualmente en una máquina de estados finitos. Esta máquina contiene la información de las posibles secuencias de caracteres que puede conformar cualquier *token* que sea parte del lenguaje (las instancias individuales de estas secuencias de caracteres son denominados lexemas). Por ejemplo, un *token* de naturaleza *entero* puede contener cualquier secuencia de caracteres numéricos.

Su principal función consiste en leer los caracteres de entrada y elaborar como salida una secuencia de componentes léxicos que utiliza el analizador sintáctico para hacer el análisis. Esta interacción, suele aplicarse convirtiendo al analizador léxico en una subrutina o corrutina del analizador sintáctico. Recibida la orden "Dame el siguiente componente léxico" del analizador sintáctico, el analizador léxico lee los caracteres de entrada hasta que pueda identificar el siguiente componente léxico. Otras funciones que realizan se listan a continuación:

- Eliminar los comentarios del programa.
- Eliminar espacios en blanco, tabuladores, retorno de carro, etc, y en general, todo aquello que carezca de significado según la sintaxis del lenguaje.
- Reconocer los identificadores de usuario, números, palabras reservadas del lenguaje y tratarlos correctamente con respecto a la tabla de símbolos (solo en los casos que debe de tratar con la tabla de símbolos).
- Llevar la cuenta del número de línea por la que va leyendo, por si se produce algún error, dar información sobre donde se ha producido.
- Avisar de errores léxicos. Por ejemplo, si @ no pertenece al lenguaje, avisar de un error.
- Puede hacer funciones de preprocesador.

Un **token** es una cadena con un significado asignado y, por lo tanto, identificado. Está estructurado como un par que consta de un «nombre de token» y un «valor de token» opcional.

El nombre del token es una categoría de unidad léxica. Los nombres de token comunes son:

- identificador: nombres que elige el programador;
- palabra clave: nombres que ya están en el lenguaje de programación;
- separador (también conocidos como signos de puntuación): caracteres de puntuación y delimitadores emparejados;
- operador: símbolos que operan sobre argumentos y producen resultados;
- literal: literales numéricos, lógicos, textuales, de referencia;
- comentario: línea, bloque (Depende del compilador si el compilador implementa comentarios como tokens, de lo contrario, se eliminará).

• Considere esta expresión en el lenguaje de programación C:

$x = a + b * 2;$

El análisis léxico de esta expresión produce la siguiente secuencia de tokens:

[(identificador, x), (operador, =), (identificador, a), (operador, +), (identificador, b), (operador, *), (literal, 2), (separador, ;)]

- Un nombre simbólico es lo que podría denominarse una categoría gramatical en lingüística.

Entre las razones por las cuales es necesario separar el analizador léxico del sintáctico se pueden mencionar:

- Un diseño sencillo es quizás la consideración más importante. Separar el análisis léxico del análisis sintáctico a menudo permite simplificar una u otra de dichas fases. El analizador léxico nos permite simplificar el analizador sintáctico.
- Se mejora la eficiencia del compilador.
- Un analizador léxico independiente permite construir un procesador especializado y potencialmente más eficiente para esa función. Gran parte del tiempo se consume en leer el programa fuente y dividirlo en componentes léxicos. Con técnicas especializadas de manejo de buffers para la lectura de caracteres de entrada y procesamiento de componentes léxicos se puede mejorar significativamente el rendimiento de un compilador.

Referencias:

- Aho, Alfred V.; Sethi, Ravi; Ullman, Jeffrey D. (1998). *Compiladores: Principios, técnicas y herramientas*. Traductores: Pedro Flores Suárez, Pere Botella i López (ilustrada edición). Pearson Educación. p. 820. ISBN 9684443331.
- *Práctica 1: Análisis léxico y sintáctico*. (s. f.). <http://www.it.uc3m.es/luis/pfat/p1/>