



**Universidad de Guadalajara**  
**Centro Universitario de Ciencias Exactas e Ingenierías**

**¿Qué es es un analizador léxico ?**

Alumno: Espinoza Sucilla Samuel

Carrera: Ingeniería en computación

Materia: Seminario Traductores de Lenguaje II

Maestro: Lopez Franco Michel Emanuel

Sección: D07 Martes y Jueves 5:00 PM a 7:00 PM.

Código: 214017739

Fecha: 21/08/2023

## ¿Qué es es un analizador léxico?

Un identificador léxico, también conocido como token léxico, es una unidad básica de un lenguaje de programación o un lenguaje de marcado. En el contexto de la programación, se refiere a las palabras clave, variables, nombres de funciones y otros elementos que tienen un significado especial en el lenguaje y son reconocidos por el analizador léxico o lexer.

El proceso de análisis léxico implica descomponer el código fuente en una secuencia de identificadores léxicos o tokens, que son fragmentos indivisibles con significados específicos. Estos tokens luego son procesados por el compilador o intérprete para generar la estructura y la lógica del programa.

Por ejemplo, en el lenguaje de programación C, los identificadores léxicos incluyen palabras clave como "if", "while", "int", "return", así como nombres de variables y funciones definidas por el usuario. En HTML, los identificadores léxicos son las etiquetas como "<html>", "<head>", "<body>", etc.

En resumen, un identificador léxico es una unidad básica de un lenguaje de programación o de marcado, y es el primer paso en el proceso de traducción y ejecución de un programa.

Existen reglas establecidas para los analizadores léxicos, también conocidos como lexers. Estas reglas determinan cómo se deben identificar y clasificar los tokens dentro del código fuente. Las reglas típicamente se definen utilizando expresiones regulares, que son patrones de texto que describen conjuntos de cadenas. Aquí hay algunos aspectos clave en las reglas de los analizadores léxicos:

**Expresiones Regulares:** Las expresiones regulares son el medio principal para definir las reglas de análisis léxico. Cada expresión regular describe un patrón de caracteres que corresponde a un tipo específico de token, como identificadores, números, cadenas, operadores, etc.

**Prioridad:** Si dos o más expresiones regulares coinciden con una misma cadena de caracteres, la expresión regular con la prioridad más alta se utilizará para generar el token correspondiente. Por ejemplo, si una cadena podría ser tanto un número como una palabra clave, la palabra clave tendría prioridad.

**Ignorar Espacios en Blanco:** Los analizadores léxicos suelen ignorar los espacios en blanco y los comentarios, ya que no son relevantes para la estructura del programa. Sin embargo, a veces se pueden contar los espacios en blanco para llevar un seguimiento de la información de la ubicación (para informes de errores, por ejemplo).

**Tokens Especiales:** Algunos tokens pueden requerir un tratamiento especial. Por ejemplo, las cadenas pueden contener caracteres especiales escapados, como en las secuencias de escape de caracteres (\n para nueva línea, \t para tabulación, etc.).

Acciones Semánticas: Además de simplemente reconocer los tokens, los analizadores léxicos pueden realizar acciones semánticas asociadas a cada token. Esto podría incluir la conversión de cadenas numéricas en valores numéricos reales o la construcción de una tabla de símbolos para mantener un registro de las variables y sus tipos.

En conjunto, estas reglas forman la base del proceso de análisis léxico, que es el primer paso en la compilación o interpretación de un programa. El analizador léxico descompone el código fuente en una secuencia de tokens, que luego se pasan al analizador sintáctico para construir la estructura gramatical del programa.