

Funciónes del Analizador Léxico

Un analizador léxico (también conocido como "analizador de símbolos" o "scanner") es una parte de un compilador o interprete que se encarga de analizar el código fuente y dividirlo en tokens.

Las funciones principales de un analizador léxico son:

- 1. Leer el código fuente carácter por carácter y reconocer patrones de símbolos.
- 2. Identificar los tokens (palabras clave, identificadores, números, etc.) y devolverlos al analizador sintáctico.
- 3. Identificar y reportar errores léxicos, como caracteres ilegales o cadenas no terminadas.
- 4. Tratar con comentarios y espacios en blanco, ignorándolos o manejándolos de forma adecuada.
- 5. Puede preprocesar ciertos elementos del código fuente, como el uso de directivas de preprocesador o los caracteres de escape en una cadena.

Componentes Léxicos

Los componentes léxicos son los elementos básicos del lenguaje de programación que el analizador léxico identifica y analiza. Algunos ejemplos de componentes léxicos incluyen:

- ❖ Palabras clave: son las palabras reservadas en un lenguaje de programación que tienen un significado específico.
- ❖ Identificadores: Los identificadores pueden ser cualquier combinación de letras, números y caracteres especiales.
- ❖ Constantes: son valores que no pueden ser modificados en el código. Pueden ser numéricos (como los enteros y los flotantes) o de caracteres (como las cadenas).
- ❖ Operadores: son los símbolos especiales que indican una acción a realizar en el código. Por ejemplo, los operadores aritméticos son "+", "-", "*", "/" y "=".
- ❖ Separadores: son los símbolos especiales que indican el fin de un componente léxico o la estructura del código. Por ejemplo, los separadores son los puntos y comas (;), los paréntesis () y las llaves {}
- Comentarios: son las líneas de texto que no son interpretadas por el compilador o intérprete. Sirven para agregar comentarios o documentación al código fuente.

Cada lenguaje de programación tiene sus propios componentes léxicos y su propia sintaxis, y el analizador léxico debe ser capaz de reconocerlos y analizarlos correctamente.



Los patrones son las reglas o expresiones regulares que el analizador léxico utiliza para identificar los componentes léxicos en el código fuente. Cada patrón se asocia con un componente léxico específico, y el analizador léxico busca estos patrones en el código fuente para determinar qué componentes léxicos hay en el código.

Los patrones también pueden ser especificados mediante un autómata finito, que es una representación matemática de un proceso de reconocimiento de patrones.

En resumen, los patrones son las reglas que el analizador léxico utiliza para identificar los componentes léxicos del código fuente, son especificados utilizando expresiones regulares o autómatas finitos y se asocian con cada componente léxico específico.

Lexema

Un lexema es el valor específico de un componente léxico.

Por ejemplo, el lexema de un identificador podría ser "x" o "contador", y el lexema de un número podría ser "5" o "3.14". El lexema es el valor que se guarda en el token generado por el analizador léxico.

Los lexemas también pueden ser utilizados en la generación de tablas de símbolos, que es una estructura de datos utilizada para almacenar información sobre los componentes léxicos en un programa. Estas tablas pueden incluir información como el tipo de una variable, el valor de una constante, o la dirección de memoria de una función.

En resumen, un lexema es el valor específico de un componente léxico, se guarda en el token generado por el analizador léxico, es utilizado por el analizador sintáctico para verificar la sintaxis del código y generar el código objeto y también pueden ser utilizado en la generación de tablas de símbolos.

