

Escuela Politécnica Nacional

Nombre: Fernando Elieco Huilca Chullagomez

Fecha:

Análisis léxico.

El análisis léxico es la primera fase del compilador y se encarga de transformar el flujo de caracteres del programa fuente en una secuencia de tokens que serán utilizados por el analizador sintáctico.

Su correcta implementación impacta significativamente en la eficiencia, claridad y portabilidad del compilador.

Funciones del analizador léxico.

- Eliminación de espacios en blanco y comentarios - Esto simplifica el trabajo del analizador sintáctico.
- Producción de tokens: Cada lexema (cadena concreta del código fuente) se traduce un token, que consiste en un tipo y un posible atributo.
- Manejo de errores léxicos: Detecta y trata errores como identificadores inválidos.
- Colaboración con la tabla de símbolos: Por ejemplo, al encontrar un identificador, el analizador puede buscarlo e insertarlo en la tabla.

Tokens, patrones y lexemas.

Se traducen los tres conceptos fundamentales.

- Token: unidad léxica, como if, id, num, que el analizador sintáctico
- Patrón: Descripción formal del conjunto de lexemas que representan un token (por ejemplo una expresión regular)

lexema: instancia concreta del programa fuente que coincide con un patrón.

El analizador léxico también puede asociar atributos a los tokens para que las fases posteriores puedan operar con mayor precisión.

Uso de búferes de entrada.

Para mejorar la eficiencia de la lectura, se utiliza un esquema de doble búfer. Cada búfer tiene un tamaño fijo (como un bloque de disco) y se llena alternadamente. Esto permite leer bloques completos de caracteres en lugar de uno por uno, reduciendo el número de llamadas al sistema. Además se emplean centinelas (caracteres especiales como `\0`) al final de los búferes para evitar verificaciones constantes de fin de bloque.

Errores léxicos y recuperación:

El analizador debe manejar casos donde no se puede identificar un token válido. Estrategias comunes incluyen:

- Modo de pánico: descartar caracteres hasta encontrar un token válido.
- Corrección Simple: insertar, eliminar o sustituir caracteres de forma de una coincidencia válida.
- Transformaciones más complejas: aunque posible, son costosas y poco prácticas.

Jonathan Pulgar