



INSTITUTO TECNOLÓGICO SUPERIOR DE CHICONTEPEC

INGENIERIA EN SISTEMAS COMPUTACIONALES

MATERIA: Lenguajes y Autómata

NOMBRE: Antonio Martínez Hernández

No. CONTROL: 1717V0049

UNIDAD 5 ANALISIS LEXICO

DOCENTE: Ing. Efrén Flores Cruz

Unidad 5

ANÁLISIS LEXICO

5.1 Fundamentos del analizador lexico

Un analizador lexico aisla al analizador sintactico de la presentación de lexemas de los Componentes lexicos. El analizador lexico opera bajo petición del analizador sintactico de volviendo un componente lexico conforme el analizador sintactico lo va necesitando para avanzar en la gramática. Los Componentes lexicos son los símbolos terminales de la gramática.

Definiciones:

Tokens o Símbolos terminales de una gramática.

- Identificadores, palabras reservadas, operadores

- Varios Símbolos pueden formar el mismo token

Atributos o Información adicional que tiene el token, de utilidad para el análisis sintactico y semántico

Componentes lexicos (tokens) o Unidad mínima de información que significa algo a la hora de compilar; Concepto de palabra; las frases de un lenguaje constan de Componentes.

Lexema o Una Secuencia de caracteres de entrada que comprenda en solo Componente lexico se llama lexema

Cadena de caracteres que extrae el Componente abstracto del Componente lexico.

Patrón o Descripción de la forma que han de tener las lexemas para ajustarse a un Componente.

5.2 Componentes lexicos, patrones y lexemas

Componente lexico (token)

Son las unidades lógicas que genera el analizador lexico. Formar caracteres en tokens es muy parecido a formar palabras en un lenguaje natural.

Es el conjunto de cadenas de código que produce como salida el mismo componente léxico. Cada token es una secuencia de caracteres que representan una unidad de información en el programa fuente. Los componentes léxicos más comunes son los siguientes palabras clave o reservadas

- | | |
|--------------------------|------------------------|
| - operadores aritméticos | - Identificadores |
| - Operadores | - Constantes |
| - Palabras clave | - Cadenas |
| - Operadores lógicos | - Literales |
| - Operador de asignación | - Signos de puntuación |

S.3 Creación de tabla de Tokens

Tabla: Conjunto de pares Clave-Valor, llamados elementos de la tabla. La tabla de símbolos es una componente necesaria de un compilador. Al detectar un identificador este es insertado en la tabla. Cada vez que se utilice el identificador se realizará una búsqueda en la tabla para obtener la información asociada

Búsqueda: dada la clave un elemento, encontrar su valor

Insertación: Dado un par clave-valor, añadir un elemento nuevo a la tabla

Cambio de Valor: Buscar el elemento y cambiar su valor

Borrado: Eliminar un elemento de la tabla

Longitud de búsqueda (o tiempo de acceso)

5.4 Errores lexicos

El analisis lexico constituye la primera fase, aqui se lee el programa fuente de izquierda a derecha y se agrupan en Componentes lexicos (tokens), que son secuencias de caracteres que tienen un significado. Ademas, todos los espacios en blanco, lineas en blanco, comentarios y demas informacion innecesaria se elimina del programa fuente. Tambien se comprueba que los simbolos del lenguaje se han escrito correctamente.

Como la tarea que realiza el analizador lexico es un caso especial de coincidencia de patrones, se necesitan los metodos de especificacion y reconocimientos de patrones y estos metodos son principalmente las expresiones regulares y los automatas finitos. Sin embargo, un analizador lexico tambien es la parte del traductor que maneja la entrada del codigo fuente y puesto que esta entrada a menudo involucra un importante gasto a tiempo, el analizador lexico debe funcionar de manera tan eficiente como sea posible.

Son pocas las errores simplemente en el nivel lexico ya que tiene una vision muy restringida de un programa fuente. El analizador lexico debe devolver el componente lexico de un identificador y dejar a otra fase se ocupa de los errores.

El compilador tiene que:

Reportar clara y exactamente la presencia de errores

Recuperarse de cada error lo suficientemente rapido para poder detectar errores subsiguientes.

Tratar de evitar mensajes falsos de error, un error que produce un token erroneo, Errores lexicos posibles

Un token o Componente léxico es una cadena de caracteres que tiene un significado coherente en cierto lenguaje de programación. Ejemplo de tokens, podrían ser palabras claves (if, while, int) identificadores, números, signos, o un operador de varias caracteres. Son los elementos más básicos sobre los cuales se desarrolla toda traducción de un programa, siendo en la primera fase llamada análisis léxico.

5.5 Generadores de analizadores léxicos

Se pueden usar varias técnicas para reducir el algoritmo y ahorrar espacio. Las etiquetas pueden guardarse mediante dispositivos para producir un sistema de finas que reduce los búsquedas.

Como las tablas de transiciones son más escasas, se pueden guardar en una lista corta que se consulte cada vez que se necesite hacer una transición a partir de un estado.

Estas listas no suelen tener más de tres o cuatro elementos, a si que su búsqueda sería por normalmente rápida.