



Seminario de Solución de Problemas de Traductores de Lenguajes II

Sección D02

Brandon Arturo Martínez González
214554831

Analizador Léxico

EL análisis léxico es la primera etapa de un compilador. El analizador léxico (o de escaneo) es un módulo destinado a leer el flujo de caracteres que componen el programa fuente, o un archivo de entrada en el cual se encuentra la cadena a analizar, y los agrupa en secuencias significativas (*lexemas*). Para cada lexema, el analizador léxico produce un *token* de la forma:

(nombre-token, valor-atributo)

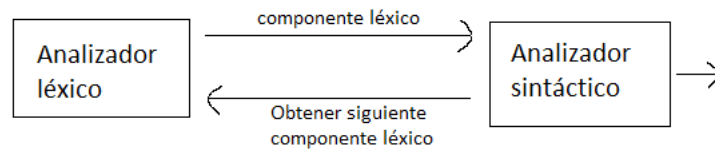
que pasa a la siguiente fase.

- Nombre-token es un símbolo abstracto que se utiliza durante el análisis sintáctico.
- Valor-atributo apunta a una entrada en la tabla de símbolos para ese token, que se necesitará para el análisis semántico y la generación de código.

El analizador léxico reconoce las palabras en función de una gramática regular de manera que sus no terminales se convierten en los elementos de entrada de las fases posteriores. El analizador léxico ignora los espacios en blanco que separan a los lexemas.

Existe una gran variedad de generadores de analizadores léxicos. Quizás la herramienta más conocida es *Lex*, un generador de analizadores léxicos para el sistema operativo Unix basada en expresiones regulares que generan código C. La organización básica del analizador léxico resultante es un autómata finito.

La interacción entre el analizador léxico y el analizador sintáctico se suele representar de la siguiente forma:



donde el analizador léxico se convierte en una subrutina del analizador sintáctico. Recibida la orden "obtener siguiente componente léxico" del analizador sintáctico, el analizador léxico lee los caracteres de entrada hasta que pueda identificar el siguiente componente léxico.