



Universidad de Guadalajara

Centro Universitario de Ciencias
Exactas e Ingeniería
2021A

Actividad: ¿Qué es un analizador léxico?

Nombre: María Fernanda Quintero Zúñiga

Materia: S.S.P de Traductores de Lenguaje II

Maestro: Michel Emanuel López Franco

¿Qué es un analizador léxico?

El Análisis Léxico es la primera fase de un compilador y consiste en un programa que recibe como entrada el código fuente de otro programa (secuencia de caracteres) y produce una salida compuesta de tokens (componentes léxicos) o símbolos. Estos tokens sirven para una posterior etapa del proceso de traducción, siendo la entrada para el Análisis Sintáctico.

La especificación de un lenguaje de programación a menudo incluye un conjunto de reglas que definen el léxico. Estas reglas consisten comúnmente en expresiones regulares que indican el conjunto de posibles secuencias de caracteres que definen un token o lexema.

En algunos lenguajes de programación es necesario establecer patrones para caracteres especiales (como el espacio en blanco) que la gramática pueda reconocer sin que constituya un token en sí. [1]

Función del analizador léxico

Su principal función consiste en leer los caracteres de entrada y elaborar como salida una secuencia de componentes léxicos que utiliza el analizador sintáctico para hacer el análisis. En la figura 1 se puede apreciar el esquema de una interacción que se aplica convirtiendo el analizador léxico en una subrutina o corrutina del analizador sintáctico. Recibida la orden "obtener el siguiente componente léxico" del analizador sintáctico, el analizador léxico lee los caracteres de entrada hasta que pueda identificar el siguiente componente léxico. [1]

Conceptos básicos del analizador léxico

Para entender el analizador léxico debemos conocer una serie de conceptos y sus definiciones

Token: Como hemos visto anteriormente, es una agrupación de caracteres reconocidos por el analizador léxico que constituyen los símbolos con los que se forman las sentencias del lenguaje y también se les denomina componentes léxicos. Constituyen los símbolos terminales de la gramática:

Palabras reservadas. Identificadores. Operadores y constantes. Símbolos de puntuación y especiales.

Lexema: Es la secuencia de caracteres, ya agrupados, que coinciden con un determinado token, como por ejemplo el nombre de un identificador, o el valor de un número. Un token puede tener uno o infinitos lexemas. Por ejemplo, las palabras reservadas tienen un solo lexema, mientras que los identificadores o los números tienen infinitos

Patrón: Es la forma de describir los tipos de lexema. Esto se realiza utilizando expresiones regulares. [2]

Errores léxico

Los errores léxicos son detectados, cuando durante el proceso de reconocimiento de caracteres, los símbolos que tenemos en la entrada no concuerdan con ningún patrón. Hay que tener en cuenta que hay pocos errores detectables por el analizador léxico, entre ellos están:

Nombres incorrectos de los identificadores: Se debe a que se utilizan caracteres inválidos para ese patrón, como por ejemplo un paréntesis, o se empieza por un número.

Números incorrectos: Debido a que está escrito con caracteres inválidos (puntos en lugar de comas) o no está escrito correctamente.

Palabras reservadas escritas incorrectamente: Se producen errores de ortografía. El problema aquí es cómo distinguir entre un identificador y una variable reservada.

Caracteres que no pertenecen al alfabeto del lenguaje: Ejemplos: @, €, ¿, ñ, etc. [2]

Bibliografía

- [1] E. Silva, «sites,» 27 Junio 2015. [En línea]. Available:
<https://sites.google.com/site/compiladoresesilval/home/compiladores/analisis-lexico>. [Último acceso: 18 Agosto 2021].
- [2] U. E. d. Madrid, «ANÁLISIS LÉXICO / CONCEPTOS BÁSICOS DEL ANÁLISIS LÉXICO,» Laureate International Universities, Madrid.