

Manual de Usuario

Analizador Léxico

Es la primera fase de un compilador, consistente 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 analizador sintáctico (en inglés parser).

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.

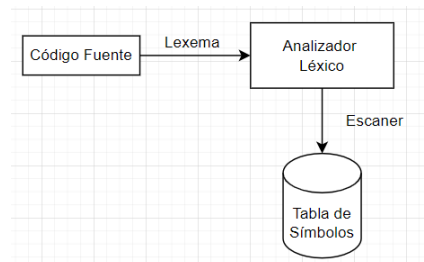
Análisis

Esta etapa está basada usualmente en una máquina de estados finitos. Esta máquina contiene la información de las posibles secuencias de caracteres que puede conformar cualquier token que sea parte del lenguaje (las instancias individuales de estas secuencias de caracteres son denominados lexemas). Por ejemplo, un token de naturaleza entero puede contener cualquier secuencia de caracteres numéricos.

Token

Un token léxico o simplemente token es una cadena con un significado asignado y, por lo tanto, identificado. Está estructurado como un par que consta de un «nombre de token» y un «valor de token» opcional. El nombre del token es una categoría de unidad léxica.¹ Los nombres de token comunes son

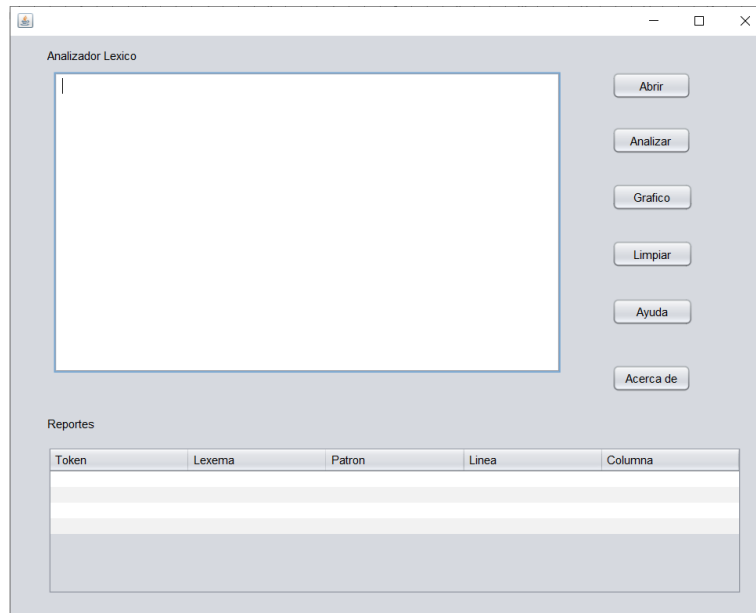
- Identificador: nombres que elige el programador;
- Palabra clave: nombres que ya están en el lenguaje de programación;
- Separador (también conocidos como signos de puntuación): caracteres de puntuación y delimitadores emparejados;
- Operador: símbolos que operan sobre argumentos y producen resultados;
- Literal: literales numéricos, lógicos, textuales, de referencia;
- Comentario: línea, bloque (Depende del compilador si el compilador implementa comentarios como tokens, de lo contrario, se eliminará).



Aquí tienes las instrucciones de cómo utilizar el programa AL conocido por Analizador Léxico basado en Python.

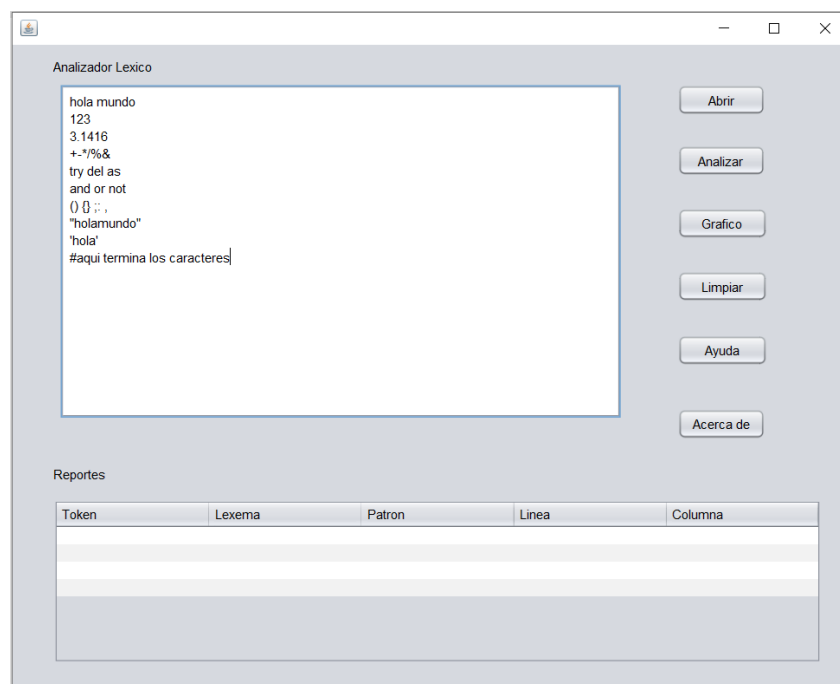
Inicio del Programa AL

Doble click para iniciar el programa, te daba una imagen como es el programa.



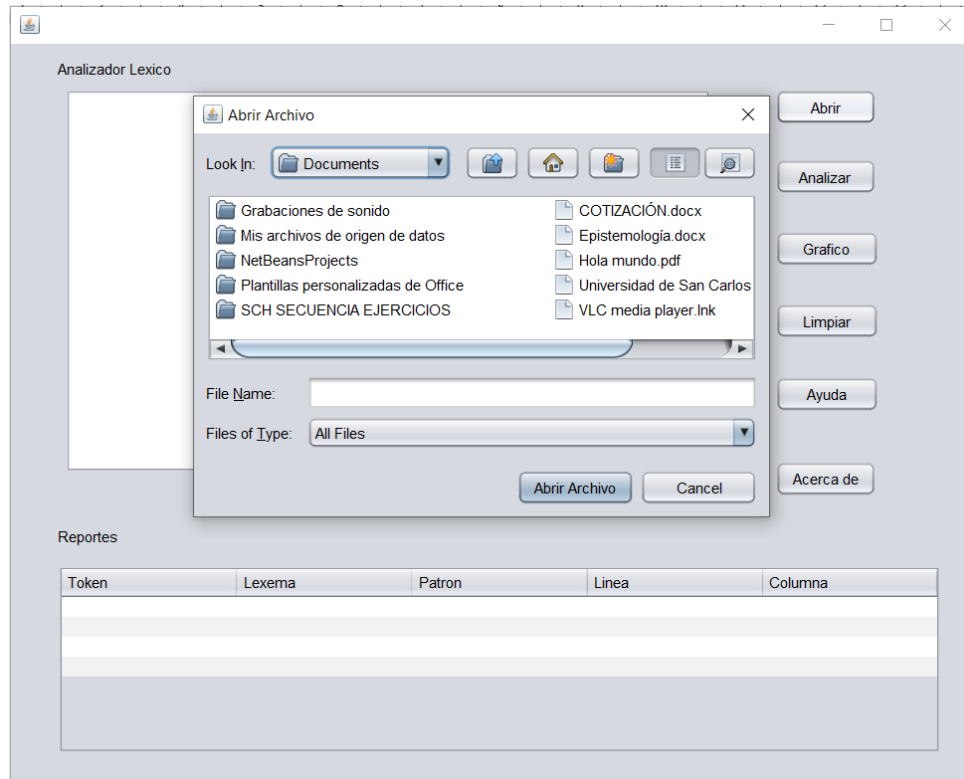
Insertar código

Puedes escribir código Python, recuerda que es un analizador léxico así que no importa la estructura que uses solo va analizar los caracteres.



Abrir archivo txt

Puedes abrir un archivo txt con el código Python, solo debes localizar el archivo (solo se permite los archivos de fuente txt).



Botón Analizar

Presiona el botón analizar y abajo observarás una tabla con todos los caracteres ingresados, te muestra los siguientes datos ingresados (el tipo Token, Lexema la palabra ingresada, Patrón basado en expresiones regulares, la Línea donde está localizado el carácter, la Columna donde está localizado el carácter).

Puedes observar que los caracteres cambian de color según su token:

Identificador color negro.

Operadores aritméticos, comparación, lógicos, asignación de color celeste.

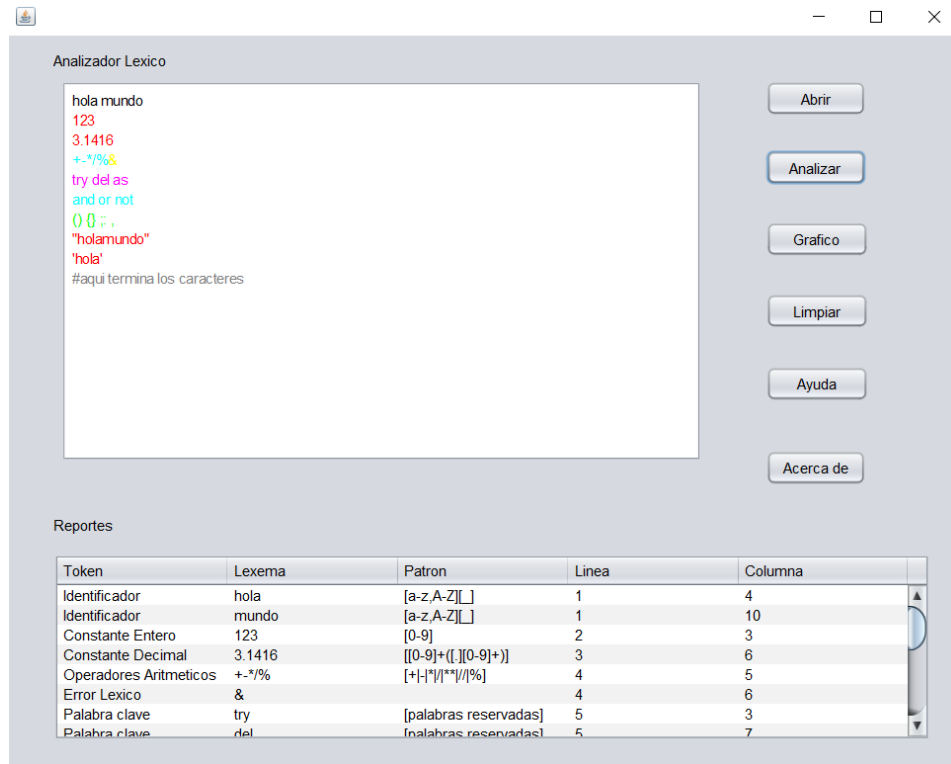
Palabras clave color morado.

Constantes color rojo.

Comentarios color gris.

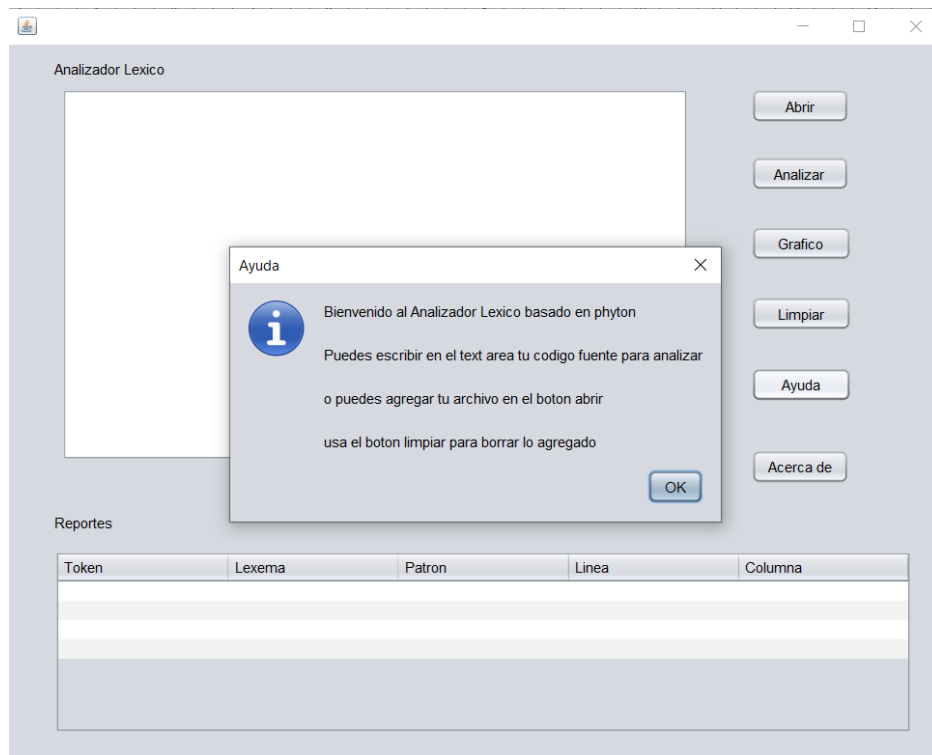
Otros color verde.

Errores léxicos color amarillo.



Botón Ayuda

Una ayuda de que se trata el programa y como puede comprenderlo.



Botón Acerca de

Datos quien realizo el programa.

