

Primer programa
Analizador Léxico**Objetivo**

Elaborar un analizador léxico en *lex/flex* que reconozca los componentes léxicos pertenecientes a las clases abajo descritas.

Descripción

- Las clases de los componentes léxicos válidos para el analizador léxico son:

Clase	Descripción
0	Identificadores. Pueden contener letras mayúsculas y minúsculas, dígitos y guion bajo. Puede empezar con letras o con dígitos. En este último caso deberá estar seguido por letra o guion bajo. Ejemplos: 1identificador, A, identificador, 1_, 16_a3, a_13_ident, 2a.
1	Constantes numéricas enteras. En base 10 (secuencia de dígitos del 0-9 sin iniciar con 0), en base 8 (inicien con 0 y le sigan dígitos del 0 al 7) y hexadecimal (inicien con 0x o 0X y le sigan al menos un dígito: 0 al 9 o 'A' a la 'F' o 'a' a la 'f')
2	Operadores relacionales (ver tabla).
3	Operadores de asignación (ver tabla)
4	Palabras reservadas (ver tabla).
5	Símbolos especiales . : { } ()
6	Operadores aritméticos (ver tabla)
7	Constantes cadenas. Iniciar y terminar con comillas dobles, cualquier carácter excepto las comillas dobles. Tamaño entre 2 y 40 (considerando las comillas).
8	Constantes numéricas reales. No signados. Casos: 2.67 .85 31. Al definir la expresión regular, no debe incluir la cadena. como constante real.

El número de clase es inamovible.

- El analizador léxico tendrá como entrada un archivo con el programa fuente, **el cual se indicará desde la línea de comandos** al momento de mandar a ejecutar el analizador léxico.
- Como delimitador de un componente léxico será uno o varios espacios, tabuladores o saltos de línea, **así como el inicio de otro componente léxico**. Considerar el orden en las acciones en el programa *lex/flex*
- Los tokens se representarán en una estructura con dos campos:
campo1: la clase
campo2: el valor (de acuerdo con las siguientes tablas e indicaciones)

Valor	Palabra reservada
0	cadena
1	devuelve
2	entero
3	flotante
4	hacer
5	mientras
6	nulo
7	para
8	si
9	sino
10	vacio

Valor	Op. relacional	significado
0	<M>	mayor que
1	<m>	menor que
2	<=>	igual que
3	<M=>	mayor o igual
4	<m=>	menor o igual
5	<\$=>	diferente

Valor	Op. aritmético	significado
0	sum	+
1	res	-
2	mult	*
3	div	/
4	mod	%
5	pow	^

Valor	Op. asignación	significado
0	i_i	igual
1	M_i	más igual
2	m_i	menos igual
3	a_i	por igual
4	d_i	entre igual
5	p_i	módulo igual
6	A_i	&=
7	P_i	^=
8	B_i	=

El valor para el token de cada identificador es la posición dentro de la tabla de símbolos. Para los operadores relacionales, los de asignación, los aritméticos y las palabras reservadas, el valor del token será la posición en su correspondiente tabla (catálogo).

Las constantes numéricas reales se incluirán en una tabla de literales, las cadenas tendrán su propia tabla de literales; por lo que el valor de su token será la posición dentro de la tabla de literales correspondiente. Cada que el Analizador Léxico encuentre una cadena o constante numérica real, NO revisará si ya se encuentra en su tabla de literales, simplemente la insertará. Para la clase 5 el valor en su token es el mismo carácter o su correspondiente ASCII

- Cuando detecte un error léxico, deberá seguir el reconocimiento a partir del siguiente símbolo.
- El analizador deberá crear la Tabla de Símbolos para almacenar a los identificadores. Esta tabla manejará los campos: posición, nombre del identificador y tipo (este último será de tipo entero y podrán ponerle como valor inicial -1). Se indicará en el documento a entregar, el tipo de estructura de datos empleada, así como el método de búsqueda a utilizar.
- Las Tablas de Literales, una para las constantes numéricas reales y otra para las cadenas, deberán tener como estructura, dos campos: la posición y el dato (cadena o constante numérica real) según el caso.
- Al término del análisis léxico deberá mostrar la tabla de símbolos, las tablas de literales, así como los tokens. También podrán almacenarse en archivos para su mejor revisión.
- Los errores que vaya encontrando el analizador léxico, los podrá ir mostrando en pantalla o escribirlos en un archivo. Deberá recuperarse de los errores encontrados para continuar con el reconocimiento de todos los componentes léxicos del archivo de entrada.
- El programa deberá estar comentado, con una descripción breve de lo que hace (puede ser el objetivo indicado en este documento), el nombre de quién(es) elaboró(aron) el programa y fecha de elaboración, así como lo que hace cada función. Se deberá cuidar mucho la sangría que denota la dependencia de instrucciones.

Entregar:

Un documento con la siguiente estructura:

- Descripción del problema (no del programa), incluyendo características que deben cumplir los componentes léxicos que reconocerá el analizador léxico y **la expresión regular de cada clase**.
- Propuesta de solución y fases del desarrollo del sistema: Análisis (planificación, indicando participantes por cada actividad), Diseño e Implementación. En el diseño indicar exactamente cómo se definirá la tabla de símbolos, las tablas de literales y los tokens, la técnica de búsqueda e inserción de los identificadores y la técnica de inserción de las cadenas y constantes numéricas reales.
- Indicaciones de cómo correr el programa.
- Conclusiones por cada participante.

Nota: se podrá elaborar individualmente o en equipo de 2

Enviar el documento y sólo el programa fuente definitivo a la plataforma educativa (apartado tareas y empaquetados en .rar o.zip) y en su caso, sólo un miembro del equipo.

Fecha de entrega: 17 de marzo de 2022.