

Primer programa  
Analizador Léxico

**Objetivo**

Elaborar un analizador léxico en *flex* o ANSI C 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 (sólo letras minúsculas)
1	Palabras reservadas
2	Operador de asignación =
3	Operadores relacionales
4	Operadores aritméticos
5	Símbolos especiales , ; [ ] ( )
6	Constantes numéricas enteras (base 10)
7	Constantes numéricas reales (siempre con .)
8	Constante cadena ( entre comillas)

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.**
- Los tokens se representarán en una estructura con dos campos:  
campo1: la clase  
campo2: el valor (de acuerdo a las siguientes tablas e indicaciones)

Valor	Op.aritméticos
0	DIVIDE
1	MAS
2	MENOS
3	MULTIPLICA

Valor	Op relac.
0	>
1	>=
2	<
3	<=
4	==
5	!=

Valor	Palabra reservada
0	ENT
1	ESCRIBE
2	HAZ
3	LEE
4	MIENTRAS
5	REAL
6	SI
7	SINO

El valor para el token de cada identificador es la posición dentro de la tabla de símbolos; de las cadenas la posición en su correspondiente tabla, y de las constantes numéricas su valor numérico (ya sea entero o real).

Para las clases 2 y 5, el valor en su token es el mismo carácter o su correspondiente ASCII

- Deberá reconocer y descartar comentarios que están definidos por iniciar con \$\$ y hasta el final de línea
- Cuando detecte un error léxico, deberá seguir el reconocimiento a partir del siguiente símbolo válido.
- El analizador deberá crear la Tabla de Símbolos para almacenar a los identificadores, los cuales estarán formado por sólo letras minúsculas. Esta tabla manejará los campos: posición, nombre del identificador y tipo (este último se quedará en blanco). Se indicará en el documento a entregar, el tipo de estructura de datos empleada así como el método de búsqueda.
- El Analizador también creará y actualizará la tabla de cadenas, la cual manejará los campos: posición y cadena. La cadena se almacenará con las comillas que la delimitan. Al igual que para los identificadores, el valor en el token de las cadenas será la posición dentro de esta tabla.
- Al término del análisis léxico deberá mostrar las tablas de símbolos y cadenas 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.
- El programa deberá estar documentado, 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 definirán las tablas de símbolos y cadenas, así como los tokens, sus técnicas de búsqueda e inserción.
- 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 y en su caso, sólo un miembro del equipo.

**Fecha de entrega: 19 de septiembre de 2017.**