## FORMATO DEL FICHERO CON LOS TOKENS GENERADOS POR EL ANALIZADOR LÉXICO

Para facilitar la depuración y la corrección de la Práctica de Procesadores de Lenguajes, y como se indica en la especificación de la misma (<a href="http://dlsiis.fi.upm.es/procesadores/Practica.html">http://dlsiis.fi.upm.es/procesadores/Practica.html</a>), el Procesador deberá leer (a través de la línea de comandos) el programa fuente de un archivo de texto y entregar obligatoriamente los resultados generados en varios archivos de texto (lista de *tokens*, Tabla de Símbolos, *parse*, errores). Se detalla aquí el formato que ha de tener el fichero donde se almacenarán los *tokens* obtenidos por el Analizador Léxico.

Debe escribirse un único *token* por cada línea. El fichero puede tener líneas vacías, que no representarán a ningún *token*. Los *tokens* tendrán el siguiente formato:

#### Dónde:

- **del**\* → cualquier cantidad de espacios en blanco o tabuladores, o nada.
- **código** → el código del *token* correspondiente, con el siguiente formato:
  - $(1 \mid d)^+ \rightarrow$  caracteres alfanuméricos, habiendo al menos uno.
- **atributo** → el atributo opcional del *token* correspondiente, que puede tener uno de los siguientes formatos:
  - nombre: (1 | d) + → caracteres alfanuméricos, habiendo al menos uno
  - número:  $[+|-]d^+ \rightarrow$  número entero con signo opcional
  - cadena: "c\*" → cadena de caracteres
- **RE** → salto de línea (RC) o Fin de Fichero (EOF)

Aunque la codificación habitual del código y del atributo del *token* es mediante un número, se permite también una representación más legible usando otros caracteres.

Tanto en el código como en el atributo del *token* no se distingue entre caracteres alfabéticos en mayúscula o en minúscula.

Se permiten comentarios precedidos por //. Un comentario puede situarse en una línea tras el *token* o en una línea que tenga sólo el comentario.

Los tokens estarán en el fichero en el orden en que son generados por el Analizador Léxico, es decir, el primer *token* del fichero será el primer *token* que haya reconocido el Analizador Léxico.

# **Ejemplos:**

Sea un Analizador Léxico que reconoce los siguientes elementos:

- Palabra reservada if
- Identificador hola
- Operador suma
- Número 99
- Cadena 'Hola, ¿qué tal?'

Ejemplos de ficheros correspondientes a dicha situación que **cumplen** el formato:



#### **Ejemplo 1:**

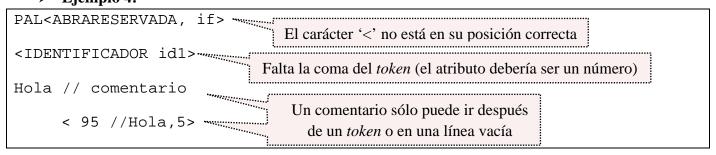
```
<1, -75> // token Palabra Reservada IF
<3, 8> // token Identificador
<57,> // token Suma sin atributo
<2,99> // token número entero
<43, "Hola, ¿qué tal?"> // token cadena
```

#### **Ejemplo 2:**

### **Ejemplo 3:**

Ejemplos de ficheros que **no cumplen** el formato:

#### > Ejemplo 4:



#### **Ejemplo 5:**

```
<PALABRA_RESERVADA, > _____ El carácter '_' no es válido dentro del código del token
<6 , ______ No se puede poner un token en dos líneas
3 > ______ Debería cumplir el formato del token o comentario
```

