



**LENGUAJES FORMALES**  
**Enunciado Práctica 1**  
**Publicación: 6 abril 2021**  
**Último día Entrega: 20 abril 2021**  
**En grupos de tres**

El lenguaje XML, al igual que HTML, define cinco **caracteres especiales** como son >, <, &, " y ', cada uno con un significado específico dentro del lenguaje. El resto de símbolos pueden ser libremente empleados para definir elementos de información. Un **elemento de información** está delimitado por un **tag de apertura** que tiene la forma <nombre> y un **tag de cierre** de la forma </nombre>, donde el primer nombre y el segundo coinciden. El **nombre de un tag** es una secuencia de caracteres alfanuméricos sin acentos, ni separadores, ni signos de puntuación (excepto la línea baja \_) que diferencian entre mayúsculas y minúsculas. Los saltos de línea tampoco son aceptados. Por ejemplo, hola, 98, a34Phj00101 y 98\_titulo son nombres válidos, mientras que hola.adios, x:y, 98\_título, nombre-autor, nota alumno (con un espacio en blanco intermedio) y hola\nadios (con \n salto de línea) son considerados nombres incorrectos. Al diferenciar entre mayúsculas y minúsculas los nombres abc, Abc, aBc, abC, ABc, AbC, aBc y ABC, por ejemplo, son considerados todos diferentes entre sí.

Un **elemento simple** es aquel elemento que se compone de un tag de apertura, seguido de un **texto** y finalizado por un tag de cierre con igual nombre que el tag de apertura. El texto intermedio puede contener cualquier carácter, incluidos los controles de salto de línea o tabuladores, pero ninguno de los cinco caracteres especiales del lenguaje XML. Si el texto intermedio no tiene ningún carácter, el elemento se denomina **elemento vacío**.

Los tags de apertura pueden contener atributos. Un **atributo** es una expresión de la forma nombre="valor" o nombre='valor', estando el nombre de los atributos sujeto a las mismas restricciones que los nombres de los tags. El signo = puede estar precedido y seguido por cero o más espacios en blanco, y el valor de un atributo puede contener un texto sujeto a las mismas restricciones que el texto entre un tag de apertura y su tag de cierre, pero sin saltos de línea. Los atributos se colocan entre nombre del tag de apertura y el cierre (símbolo >) del tag de apertura, separándose diferentes atributos con uno o más espacios en blanco.

Por ejemplo, la información de una mujer llamada María José García Lluch que vive en la calle San Francisco 28, 3º 5ª, de Tarragona y cuyo número de teléfono es desconocido, puede representarse con el lenguaje XML como:

```
<Genero gen = 'mujer'></Genero>
<nombre compuesto="cieRto">María José </nombre>
<apellido>García Lluch</apellido>
<telefono></telefono>
<direccion>Calle San Francisco nº 28, 3-5, 43010 Tarragona<direccion>
```

Se pide: Emplear la herramienta Flex para contabilizar el número de tags de apertura, el número de tags de cierre, el número de tags con uno o más atributos, el número total de atributos y la cantidad de caracteres de todos los textos que aparecen en un documento XML.

Para ello, deben definirse expresiones regulares correctas para cada uno de los siguientes conceptos: tag de apertura, tag de cierre, nombre de un tag, texto, y atributo (5 puntos). Además, debe hacerse la contabilización de todos los tags de apertura, tags de cierre, elementos simples, elementos vacíos y elementos con atributos (5 puntos).

*NOTA: la descripción del lenguaje XML que se hace en este enunciado es una simplificación del lenguaje XML real.*