

Clasificación y características comunes de los lenguajes de programación.

1.2.1 Clasificación de los lenguajes de marcas

El uso de los lenguajes de marcas es muy diverso. Se utiliza para mensajería instantánea XMPP (extensible messaging and presence protocol), servicio web WSDL (web service language), SOAP (simple object Access protocol), páginas web HTML, WML (wireless markup language), sindicato de contenidos RSS (really simple syndication) Atom, graficos vectoriales X3D, documentación electrónica RTF (rich text format)., TeX, wikitexto, etc.

Los lenguajes de marcas suelen clasificarse en tres tipos:

Tipo 1. De procedimiento. Suele emplearse para la presentación del texto. Siendo visible para el usuario. Pueden utilizarse etiquetas para poner un titulo centrado, reducir el tamaño de letra poner el texto cursivo, etc. Ejemplo de este tipo de marcado son TeX, Nroff y La TeX.

Tipo 2. De presentación. Se refiere al que define el formato del texto. Aunque fácil de crear, resulta complicado de mantener o modificar. Las etiquetas de marcado, por lo general, no son visibles a los usuarios. Unos de sus usos puede ser la maquetación de documentos para que el lector pueda leerlos correctamente. Por ejemplo: Microsoft Word.

Tipo 3. Descriptivos o semánticos. Es un marcado flexible que usa etiquetas sin especificar la manera de presentarla ni el orden. Las marcas dan información sobre estructuras y una descripción del contenido. Algunos de estos lenguajes son: SGML, HTML, XML.

En el ejemplo se observa una serie de etiquetas creadas por un documento XML que almacena información relacionada con los exámenes de distintas materias de un ciclo. En este caso concreto, se han creado dos materias de lasque se almacena el nombre, el número de alumnos aprobados y suspensos y la unidad de cada una.

```
<?xml version="1.0" encoding="iso-8891-1"?>
<exámenes>
  <materia>
    <nombre> Lenguaje de marcas </nombre>
    <aprobados> 15 </aprobados>
    <suspensos> 10 </suspensos>
    <unidades > 4 </unidades >
  </materia>
  <materia>
    <nombre> Sistemas operativos</nombre>
    <aprobados> 5 </aprobados/>
    <suspensos> 20 </suspensos>
    <unidades > 5 </unidades>
  </materia>
</exámenes>
```

1.2.2 Características comunes

No todos los lenguajes de marcas comparten las mismas características debido a la gran cantidad que pueden encontrarse en el mercado actual. En este apartado, se tratarán las características comunes más importantes de estos lenguajes como puede ser el uso de texto plano, flexibilidad, interoperabilidad.

A) Texto plano (plain text).

Este tipo de documentos, también llamado texto sin formato (color, tipo de letra, negrita, tamaño, subrayados, etc.) tal y como puede hacerlo un procesador de texto avanzado. Están formados exclusivamente por caracteres (letras, números, caracteres especiales y de control).

B) Interoperabilidad o independencia

Se considera el texto plano como formato universal, ya que puede abrirse y editarse desde cualquier máquina, aunque hay que tenerse en cuenta la codificación empleada. Son independientes a la plataforma usada y a cualquier sistema de representación.

C) Flexibles y fáciles de crear

Simplemente, se necesita un editor de texto plano para poder crearlos y guardarlos en la extensión que se desee. Algunos permiten combinarse con otro lenguaje para darle mejor funcionalidad.

1.3 XML: estructura y sintaxis

XML, sigla de extensible markup language (lenguaje de marcas extensibles), se considera un metalenguaje y surge para resolver los problemas que plantea HTML, al mostrar las etiquetas el significado de los datos. Una de las ventajas es que no requieren conocimientos de programación para poder crear o modificar un documento sencillo. La importancia radica en el intercambio de información de manera segura entre distintos programas, lo que permite la reutilización de contenido, crear etiquetas propias y presentar una estructura y diseño independientes.

Aunque este apartado se estudiara con más profundidad en el capítulo 4, se mostrará una instrucción a su estructura y sintaxis para ir asimilando la terminología empleada.

Todo documento XML tiene una estructura jerárquica arborescente y está compuesto por dos partes fundamentales:

Prologo, que será la parte del documento y está compuesto por una o varias líneas, y cuerpo.

En este ejemplo el cuerpo estará compuesto por la primera línea por donde se declara la versión XML, así como la codificación empleada para representar los caracteres, mientras que el resto del documento, el cuerpo, estaría compuesto por elementos anidados. Dentro de la etiqueta < cine >, de incluyen las distintas películas de la cartelera. En este caso, se ha introducido una película y se guarda la información relativa al director, la duración y el título.

```
<?xml version="1.0" encoding="iso-8891-1"?>
<cine >
  <película >
    <director > Señor X</director >
    <duración > 60 </duración >
    <titulo > Las aventuras de XML </titulo >
  </película >
</cine >
```

1.4 Herramientas de edición

En este apartado, se tratan las herramientas básicas relacionadas con la edición de páginas web, así como otros para la creación de documentos XML, estudiando algunos de mayor presentación en capítulos posteriores. Al igual que para otro tipo de software pueden encontrarse tantas herramientas offline, que tiene que descargarse e instalarse en el ordenador personal para poder usarlas, y online, donde el único requisito es disponer de una conexión a internet y el uso de un navegador para acceder.

1.4.1 Bloc de notas Windows

El bloc de notas Windows (Notepad) es un editor de texto que viene con el sistema operativo Windows que permite crear documentos de texto plano. La extensión de dichos documentos es .txt y permite realizar la codificación de caracteres ANSI, Unicode, UTF-8.

1.4.2 Gedit

Gedit es un editor de texto de fácil uso para el entorno de escritorio GNOME, aunque también está disponible para Windows. Suele emplearse como simple bloc de notas, utilizar las características avanzadas y usarlo como entorno de desarrollo. Además de las características generales que puede tener cualquier editor de texto como copiar, abrir, imprimir, etc., permite el coloreado del sintaxis dependiendo del lenguaje de programación usado, así como la activación de la numeración de líneas.

1.4.3 Notepad++

Es un potente editor de texto gratuito de licencia GPL para sistemas operativos Windows. Está escrito en C++ usando la API Win32. Permite seleccionar multitud de lenguajes de programación, organiza el texto según el lenguaje seleccionado y muestra entre otros aspectos, numeración de líneas, color texto en etiquetas según la sintaxis, aumento de tamaño de letra, etc.

Entre sus características, destacan las siguientes: permitir ampliar y reducir el texto, autocompletar, búsqueda y remplazo, resaltado de sintaxis.

1.5 Elaboración de documentos XML bien formados

Para poder elaborar un documento XML bien formado, es necesario seguir una serie de reglas, entre las que destacan:

- El W3C recomienda empezar por la instrucción donde se indique la versión (instrucción de procesamiento), por lo que esta ha de ser siempre la primera instrucción.
- Todo documento ha de tener una estructura jerárquica, en la que exista un único elemento raíz o principal, que será el primero que se abra y el último que se cierre.
- Las etiquetas tienen que estar anidadas correctamente unas dentro de otras, cerrándose en orden inverso al que se abren. Los elementos no vacíos tendrán etiquetas de apertura y de cierre.
- Dentro de los atributos, los valores tienen que estar encerrados entre comillas dobles o simples y no podrá haber dos atributos que se llamen igual en un mismo elemento.

- A la hora de escribir etiquetas, elementos o atributos, hay que tener en cuenta que es sensible a las mayúsculas y las minúsculas. Por ejemplo, un XML de apertura de una etiqueta <casa> y uno de cierre </Casa> son distintas, ya que una empieza con minúscula y otra con mayúsculas.
- Los nombres de etiquetas, elementos o atributos tienen que seguir una nomenclatura como no empezar por número, no usar caracteres especiales reservados, etc. Tienen que empezar con letras, seguido de guiones, puntos, números u otras letras. Tampoco pueden empezar por la palabra XML.
- Los comentarios no pueden ir dentro de las etiquetas.

```
<?xml version="1.0" encoding="iso-8891-1"?>
<agenda >
  <contacto >
    <nombre > Alumno X</nombre >
    <ciudad cp='2560' > Bata </ciudad >
    <teléfono n='2546285' / >
  </película >
</agenda >
```