1

CONCEPTOS PREVIOS

1.1 LENGUAJES DE MARCADO

Los lenguajes de marcado, también se les suele llamar lenguajes de marcas, aunque este tipo de referenciación no es nada habitual.

Históricamente, los lenguajes de marcado eran utilizados por las editoriales y medios de comunicación para imprimir instrucciones concretas en los márgenes. Para ello, utilizaban unos "marcadores" que indicaban el tipo de fuente, estilo, tamaño de letra y la corrección de errores. Una cosa llevó a la otra y, con el tiempo, se estandarizaron como un grupo de marcas que, más tarde, fueron reinterpretadas en lo que hoy conocemos como lenguajes de marcado.

El lenguaje de marcado es un término que hace referencia a una manera de codificar la información. Esta codificación se caracteriza porque los contenidos están envueltos y declarados a través de unas entidades que llamaremos etiquetas o marcas.

En este contexto, cuando se habla de etiquetas o marcas, además se les suele dar un significado que puede indicar qué es o qué debe contener, dónde utilizarlo o cómo presentarlo. No obstante, ninguna de estas marcas tiene la capacidad de ejecutar acciones ni de realizar operaciones de ninguna clase.

Esto último debe tenerse claro porque, en ocasiones, los lenguajes de marcado se confunden con los lenguajes de programación, y nada más lejos. Un lenguaje de programación es un idioma de computación que presenta varias peculiaridades como son la posibilidad de declarar variables, realizar operaciones matemáticas o crear fragmentos de código que suelen ser agrupados en procedimientos o funciones. Un lenguaje de marcado no hace nada de eso, sólo etiqueta la información aportándole un valor semántico.

1.1.1 Características de los lenguajes de marcado

Un lenguaje de marcado se caracteriza porque tiene marcas o etiquetas que son declaradas a modo de elemento o entidad. Cada uno de estos elementos o entidades tiene un contenido y uno o varios atributos que definen su comportamiento, su función y su significado.

Un ejemplo de todo esto podría ser la declaración de un producto.

```
<producto>
    <cabecera>
        <titulo>Nombre producto</titulo>
        <subtitulo>Categoría</subtitulo>
    </cabecera>

    <descripcion>
        <texto>Descripción del producto</texto>
    </descripcion>

    <pie>
        <negrita>Precio</negrita>
        <accion>Comprar</accion>
    </pie>
</producto>
```

Figura 1.1. Ejemplo de lenguaje de marcado

Como se puede observar en la ilustración anterior, se ha definido una entidad concreta a través de unos elementos que definen su significado y comportamiento y que, en ocasiones, tienen propiedades intrínsecas que aportan un valor a la presentación.

Un ejemplo de ello podría ser la etiqueta `SUBTITULO`, que indica que, su contenido textual, es un texto de cabecera o a modo de título y que está en el segundo nivel de la jerarquía en lo referente a títulos.

Si, además, este código fuese interpretado por una herramienta o sistema específico, podría llevar una interpretación específica que provocase que fuese presentado de forma concreta, con un tamaño, margen y estilo concretos, entre otras cosas.

1.1.2 Clasificación de los lenguajes de marcado

Los lenguajes de marcado suelen ser clasificados en tres grupos, no obstante, pueden llegar a mezclarse propiedades de varios en uno, formando nuevas especificaciones.

1.1.2.1 MARCADO DE PRESENTACIÓN

El marcado de presentación es aquel que está enfocado hacia la presentación de la información del documento, es decir, es aquel que sirve para dar formato al texto. Este tipo de marcado es suficiente para el proceso de lectura, pero insuficiente para el procesamiento automático de la información. Además, las etiquetas de marcado están ocultas al usuario, lo que hace que su contenido esté ofuscado y sea complicado de extraer si no es con la herramienta adecuada.

El marcado de presentación puede ser sencillo de elaborar, sin embargo, su mantenimiento y/o modificación pueden volverse complicados, por lo que su uso no está demasiado extendido.

Un ejemplo de marcado de presentación es RTF, en donde la composición de las etiquetas suele proporcionar un comportamiento y/o un significado distinto. Es decir, el título del documento podría venir predefinido por una combinación de varios saltos de línea en cualquier posición del documento y terminar con un carácter especial justo detrás del contenido textual.

```
{\rtf1\ansi\ansicpg1252\deff0\deflang3082
{\fonttbl{\f0\fswiss\fcharset0 Arial;}{\f1\fmodern\fprq1\fcharset0 Arial;}}
{\colortbl ;\red255\green0\blue0;\red0\green0\blue128;}
\viewkind4\uc1\pard\b\f0\fs20 RTF\b0 es un \cf1\b\fs24 lenguaje de marcado de
presentaci\’f3 \f0\fs20 .\par}
```

Figura 1.2. Fragmento de código de documento RTF

1.1.2.2 MARCADO DE PROCEDIMIENTOS

El marcado de procedimientos es aquel que está enfocado hacia la presentación del texto, al igual que sucede con el marcado de presentación. No obstante, las marcas o etiquetas que componen el documento no están ocultas, es decir, son visibles en todo momento para el usuario.

El marcado de procedimientos es sencillo de elaborar y de modificar. Además, es interpretado por orden de aparición, por lo que ayuda a predecir cómo será su resultado durante el proceso de lectura.

Un ejemplo de marcado de procedimientos es LaTeX y HTML. Ambos requieren de una marca o etiqueta que funciona a modo de identificador especial que le indica al sistema cómo se debe renderizar el contenido textual. Este renderizado suele venir definido a través de una serie de instrucciones que establecen el texto con un tipo de fuente, tamaño, alineación y estilos concretos.

```
<article>
    <header>
        <h2>Nombre producto</h2>
```

```
    </header>

    <div class="descripcion">
        <p>Descripción del producto</p>
    </div>

    <footer>
        <b>Precio</b>
        <button>Comprar</button>
    </footer>
</article>
```

Figura 1.3. Fragmento de código de documento HTML5

1.1.2.3 MARCADO DESCRIPTIVO

El marcado descriptivo, también conocido como semántico, es aquel que utiliza las etiquetas para describir fragmentos de texto, pero sin especificar en qué orden o cómo deben representarse.

El marcado descriptivo suele ser sencillo de elaborar y modificar y, a menudo, sus etiquetas pueden ir acompañadas de atributos que definen su significado o comportamiento, pero su interpretación está separada de la presentación, es decir, sólo especifica la descripción del tipo y contenido de los documentos.

Al marcado descriptivo también se le atribuye la cualidad de simplificar el proceso de reformateado del texto, principalmente, porque la información del formato está separada del propio contenido. Por esta razón, un fragmento indicado como cursiva a través de la etiqueta `I` (italic), podría emplearse para marcar énfasis o para indicar un contenido gráfico.

Cabe destacar que, si quisiéramos sortear esta última situación en el marcado de presentación y en el marcado de procedimientos, el proceso probablemente sería bastante tedioso, sin embargo, si los contenidos se hubiesen diferenciado descriptivamente mediante etiquetas distintas, podrían representarse de manera diferente sin esfuerzo.

Un ejemplo de marcado descriptivo es SGML y el XML. Ambos son muy flexibles ya que los fragmentos se etiquetan como son y no como deberían mostrarse.

```
<!doctype email system "email.dtd">
<email>
    <to>info@ejemplo.com
    <from>alumno@gmail.com
    <date>mon, 29 jan 2020 12:00:02 - 0100 (est)
    <subject>error de acceso
```

```
    <contents>
        Hola, buenas tardes. ¿Mi cuenta está bloqueda o deshabilitada?.
        <url>http://ejemplo.com</url>
        saludos
    </contents>
</email>
```

Figura 1.4. Fragmento de código de documento SGML