

Unidad II: Lenguaje de marcado

2.1 Introducción

Un lenguaje de marcado o lenguaje de marcas es una forma de codificar un documento que, junto con el texto, incorpora etiquetas o marcas que contienen información adicional acerca de la estructura del texto o su presentación.

El lenguaje de marcas más extendido es el HTML ("HyperText Markup Language", Lenguaje de marcado de hipertexto), fundamento del World Wide Web (entramado de comunicación de alcance mundial).

Los lenguajes de marcado suelen confundirse con lenguajes de programación. Sin embargo, no son lo mismo, ya que el lenguaje de marcado no tiene funciones aritméticas o variables, como sí poseen los lenguajes de programación. Históricamente, el marcado se usaba y se usa en la industria editorial y de la comunicación, así como entre autores, editores e impresores.

Un ejemplo de cómo funciona el lenguaje de marcado puede observarse en el dictado de viva voz de un documento a una persona que lo transcribe a máquina:

Ponga estilo de carta. Ponga comillas, ponga mayúsculas, Estimado Juan, ponga dos puntos, aparte, sangría, ponga primera letra mayúscula, te escribo esta carta, ponga negrillas, de forma muy urgente, cierre negrilla, ya que no me has enviado... etc".

2.2 Representación de documentos

- De presentación:
 - Indica el formato del texto (información para el maquetado).
- De procedimientos:
 - Orientado también a la presentación pero, en este caso, se indican los procedimientos que deberá realizar el SW de representación.
- Descriptivo o semántico:

– Describen las diferentes partes en las que se estructura el documento pero sin especificar cómo deben representarse.

2.3 Tipos de datos básicos

Recordando lo anteriormente dicho, el SGML permite definir nuevos metalenguajes. A cada lenguaje de formato de documentos definido con SGML se lo denomina aplicación SGML. Las partes en las que se divide una aplicación SGML son:

1. Una declaración SGML:

En la misma se concreta que caracteres y delimitadores pueden aparecer en la aplicación. Es decir, lo que se declara es la versión de caracteres que van a poder ser utilizados. (UTF-8 por ejemplo).

2. Una Declaración del tipo de documento (DTD):

Ésta define la sintaxis de las estructuras de formato, así como referencias a entidades de caracteres. En palabras más sencillas, define el tipo de etiquetas que van a poder ser usadas, los caracteres permitidos en el lenguaje, como se va a estructurar ese lenguaje etcétera... De esta manera se podrá luego validar frente a ese tipo de DTD seleccionada, para comprobar si la sintaxis a la hora de colorar las balizas ha sido la correcta o tiene errores, sin embargo. Validar una web frente a la DTD empleada mediante el servicio del Validador de la W3C es algo imprescindible. De otra manera el comportamiento de nuestra web frente a los navegadores podría ser imprevisible, además de que afectaría al posicionamiento web, del cual hablaremos más adelante.

3. Una especificación que describe la semántica:

Se trata de una especificación que describe el significado que se debe aportar al código de formato. Adicionalmente, esta especificación también impone restricciones de sintaxis que no pueden expresarse dentro del DTD.

4. Documentos con su contenido y su marcado entrelazado:

Cada documento, como es lógico, debe contener una referencia al DTD que debe usarlo para interpretarlo. (Validación).

Ahora vamos a ahondar aún más en todos estos conceptos:

Sobre la DTD:

2.4 Estructura global de un documento

Un documento HTML 4 se compone de tres partes:

1. una línea que contiene información sobre la versión de HTML,
2. una sección de cabecera declarativa (delimitada por el elemento HEAD),
3. un cuerpo, que contiene el contenido real del documento. El cuerpo puede ser especificado mediante el elemento BODY o mediante el elemento FRAMESET.

Puede aparecer espacio en blanco (espacios, saltos de línea, tabulaciones y comentarios) antes y después de cada sección. Las secciones 2 y 3 deberían estar delimitadas por el elemento HTML.

Aquí tenemos un ejemplo de un documento HTML sencillo:

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN"
  "http://www.w3.org/TR/html4/strict.dtd">
<HTML>
  <HEAD>
    <TITLE>Mi primer documento HTML</TITLE>
  </HEAD>
  <BODY>
    <P>¡Hola mundo!
  </BODY>
```

</HTML>

2.5 Elementos básicos: texto, vínculos, listas, tablas, objetos, imágenes y aplicaciones

Dentro del documento hay dos zonas principales: el encabezamiento, delimitado por las marcas <HEAD> y </HEAD>, que sirve para definir algunos valores válidos para todo el documento, y el cuerpo, delimitado por las etiquetas <BODY> y </BODY>, donde reside la información del documento.

El elemento <TITLE> contenido dentro del encabezamiento permite especificar el título de un documento HTML. Este título no forma parte del documento en sí pues no se ve en la pantalla principal, sino que sirve como título de la ventana del programa que la muestra.

Existen muchos otros elementos que se engloban dentro del encabezamiento pero para la estructura básica del lenguaje HTML en su nivel básico no son necesarios.

El cuerpo de un documento HTML contiene el texto, imágenes, etc. que, con la presentación y los efectos que se decidan, se presentarán ante el usuario. Dentro del cuerpo se pueden aplicar una serie de efectos a través de diferentes marcas o etiquetas (también otros autores las denominan "directivas").

Así pues, la estructura de un documento HTML es la siguiente:

```
<HTML>  
<HEAD>  
<TITLE>Título de la página</TITLE>  
</HEAD>  
<BODY>  
[Aquí se sitúan otras etiquetas que hacen posible visualizar la página]  
</BODY>  
</HTML>
```

El texto puede tener unas cabeceras, comprendidas entre las etiquetas <H1> y </H1>, <H2> y </H2>, etc. (hasta el número 6), siendo el número indicativo del tamaño. El tamaño mayor es el correspondiente al número 1.

Hay otras etiquetas como **<P>**, para separar los distintos párrafos, la etiqueta de centrado **<CENTER>** y **</CENTER>** que sirve para centrar todo lo que esté dentro de ella, ya sea texto, imágenes, etc. la etiqueta **<HR>** para obtener una raya horizontal tan ancha como la pantalla, y con la apariencia de estar embutida sobre el fondo, etc.

He aquí un ejemplo:

```
<HTML>
<HEAD>
<TITLE>Mi pagina Web</TITLE>
</HEAD>
<BODY>
<CENTER>
<H1>Primera pagina</H1>
</CENTER>
<HR>
Esta es mi primera pagina, aunque todavia es muy sencilla.
<P>Aqui va un segundo parrafo.
</BODY>
</HTML>
```

Pero existen también una serie de Atributos:

<body background="nombre_de_archivo">

Especifica una imagen para el fondo del documento.

<body bgcolor="#rrggbb">

Especifica un color para el fondo del documento.

<body text="#rrggbb">

Especifica un color para el texto normal del documento.

<body link="#rrggbb" vlink="#rrggbb" alink="#rrggbb">

Establece el color del texto de vínculo, del vínculo visitado y del vínculo activo.

#rrggbb representa el valor hexadecimal del color expresado como rojo-verde-azul.

Estilos y efectos básicos

Todas las etiquetas que siguen a continuación se introducen a partir de la etiqueta **<BODY>**, es decir, dentro del cuerpo del documento.

Títulos

Mediante los títulos, en sus diferentes niveles de importancia, se puede definir el esqueleto del documento o estructura básica. HTML tiene 6 niveles de cabeceras numeradas del uno al seis. He aquí algunos ejemplos de cabeceras HTML y su representación en pantalla:

<i>Código HTML</i>	<i>Visualización</i>
<code><h1>Cabecera tipo 1</h1></code>	Cabecera tipo 1
<code><h2>Cabecera tipo 2</h2></code>	Cabecera tipo 2
<code><h3>Cabecera tipo 3</h3></code>	Cabecera tipo 3
<code><h4>Cabecera tipo 4</h4></code>	Cabecera tipo 4
<code><h5>Cabecera tipo 5</h5></code>	Cabecera tipo 5
<code><h6>Cabecera tipo 6</h6></code>	Cabecera tipo 6

Elementos de texto

`<p>`

El fin de un párrafo que será formateado antes de que se muestre en la pantalla

`<pre>... </pre>`

Identifica texto que ha sido formateado previamente (preformateado) por algún otro sistema y debe ser desplegado tal como es. Texto preformateado puede incluir etiquetas embebidas, pero no todos los tipos de etiquetas están permitidos. La etiqueta **<tag>** puede ser usada para incluir tablas en documentos.

`<listing>... </listing>`

Ejemplo de listado de ordenador; las etiquetas incluidas serán ignoradas, pero los espacios de tabulación funcionarán. Esta es una etiqueta ya arcaica.

`<xmp>... </xmp>`

Similar a `<pre>` excepto que no se reconocerán etiquetas incluidas.

`<plaintext>`

Similar a `<pre>` excepto que no se reconocerán etiquetas incluidas, como no hay etiqueta para finalizar, el resto del documento será traducido como

texto simple. Esta es una etiqueta arcaica. Algunos navegadores reconocen una etiqueta `</plaintext>`, aún cuando no está definido por el estándar.

<blockquote> . . . </blockquote>

Incluye una sección de texto citado de cualquier otra fuente.

Atributos del texto

Mediante estos atributos determinamos el estilo y el tipo de letra que tendrá la presentación del documento final.

Estilos lógicos	Estilos físicos
<p><code> . . . </code> Énfasis</p>	
<p><code> . . . </code> Énfasis mayor</p>	
<p><code><code> . . . </code></code> Despliega una directiva HTML</p>	
<p><code><samp> . . . </samp></code> Incluye salida ejemplo</p>	<p><code> . . . </code> Negritas</p>
<p><code><kbd> . . . </kbd></code> Despliega una tecla del teclado</p>	<p><code><i> . . . </i></code> <i>Itálicas</i></p>
<p><code><var> . . . </var></code> Define una variable</p>	<p><code><u> . . . </u></code> <u>Subrayado</u></p>
<p><code><dfn> . . . </dfn></code> Despliega una definición (en muchos navegadores no funciona)</p>	<p><code><tt> . . . </tt></code> Tipo de letra de máquina de escribir</p>
<p><code><cite> . . . </cite></code></p>	

```
<FONT  
SIZE="+3">This is  
+3</FONT>
```

```
<FONT SIZE="+1">This is  
+1</FONT>
```

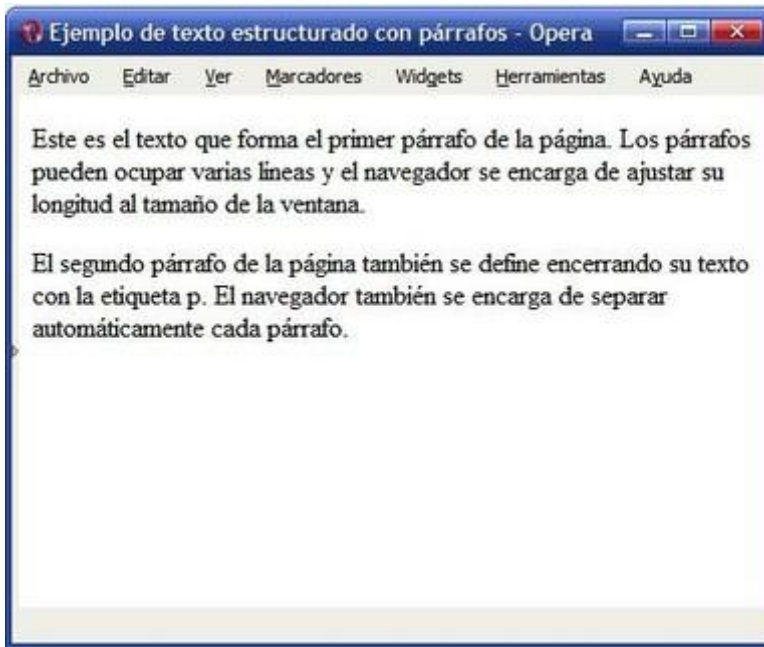
La forma más sencilla de estructurar un texto consiste en separarlo por párrafos. Además, HTML permite incluir títulos que delimitan cada una de las secciones.

Párrafos

Una de las etiquetas más utilizadas de HTML es la etiqueta <p>, que permite definir los párrafos que forman el texto de una página. Para delimitar el texto de un párrafo, se encierra ese texto con la etiqueta <p>, como muestra el siguiente ejemplo:

```
<html>  
<head>  
<title>Ejemplo de texto estructurado con párrafos</title>  
</head>  
<body>  
<p>Este es el texto que forma el primer párrafo de la página.  
Los párrafos pueden ocupar varias líneas y el navegador se encarga  
de ajustar su longitud al tamaño de la ventana.</p>  
<p>El segundo párrafo de la página también se define encerrando  
su texto con la etiqueta p. El navegador también se encarga de  
separar automáticamente cada párrafo.</p>  
</body>  
</html>
```

El ejemplo anterior se visualiza de la siguiente manera en cualquier navegador:



La siguiente tabla recoge la definición formal de la etiqueta <p>:

Etiqueta	<p>
Atributos comunes	básicos, i18n y eventos
Atributos propios	-
Tipo de elemento	Bloque
Descripción	Delimita el contenido de un párrafo de texto

Los párrafos creados con HTML son elementos de bloque, por lo que siempre ocupan toda la anchura de la ventana del navegador. Además, no tienen atributos específicos, pero sí que se les pueden asignar los atributos comunes de HTML básicos, de internacionalización y de eventos.

Secciones

Las páginas HTML habituales suelen tener una estructura más compleja que la que se puede crear solamente mediante párrafos. De hecho, es habitual que las páginas se dividan en diferentes secciones jerárquicas.

Los títulos de sección se utilizan para delimitar el comienzo de cada sección de la página. HTML permite crear secciones de hasta seis niveles de importancia. De esta forma, aunque una página puede definir cualquier número de secciones, sólo puede incluir seis niveles jerárquicos.

Las etiquetas que definen los títulos de sección son <h1>, <h2>, <h3>, <h4>, <h5> y <h6>. La etiqueta <h1> es la de mayor importancia y por tanto se utiliza para definir los titulares de la página. La importancia del resto de etiquetas es descendiente, de forma que la etiqueta <h6> es la que se utiliza para delimitar las secciones menos importantes de la página.

A continuación se muestra la definición formal de la etiqueta <h1>, siendo idéntica la definición del resto de etiquetas referidas a los títulos de sección:

Etiqueta	<h1>
Atributos comunes	básicos, i18n y eventos
Atributos propios	-
Tipo de elemento	Bloque
Descripción	Define los títulos de las secciones de mayor importancia de la página.

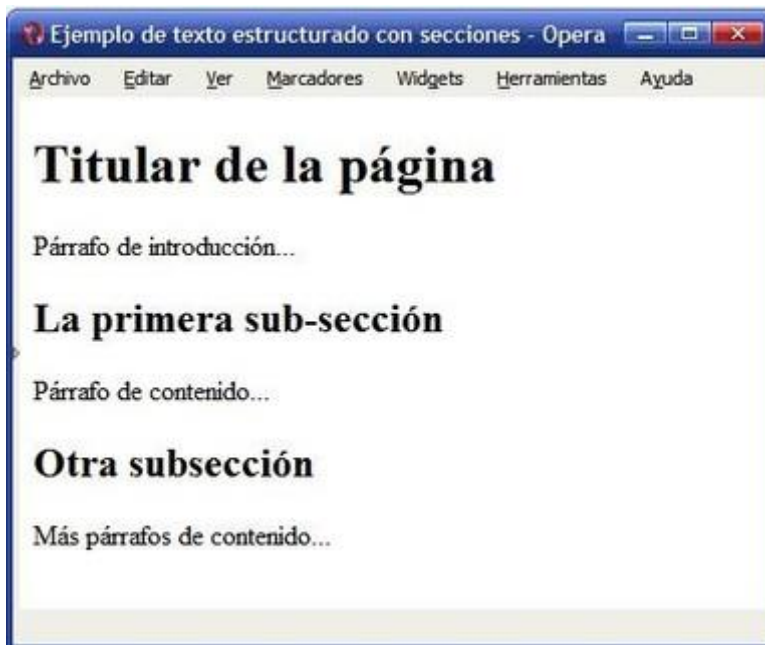
Al igual que la etiqueta <p>, las etiquetas de título de sección son elementos de bloque y no tienen atributos específicos.

Las etiquetas <h1>, ..., <h6> definen títulos de sección, no secciones completas. Por este motivo, no es necesario encerrar los contenidos de una sección con su etiqueta correspondiente. Solamente se debe encerrar con las etiquetas <h1>, ..., <h6> los títulos de cada sección.

El siguiente ejemplo muestra el uso de las etiquetas de título de sección:

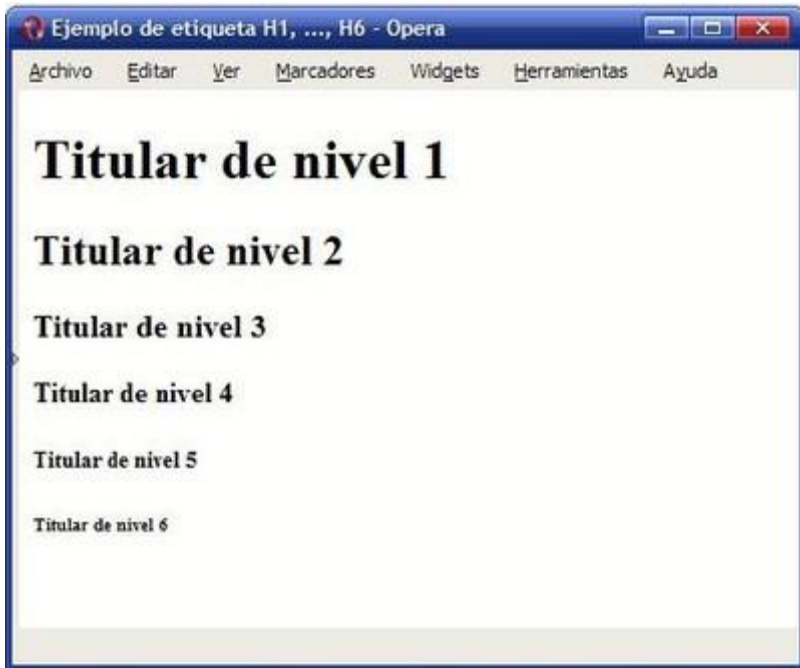
```
<html>
<head>
<title>Ejemplo de texto estructurado con secciones</title>
</head>
<body>
<h1>Titular de la página</h1>
<p>Párrafo de introducción...</p>
<h2>La primera sub-sección</h2>
<p>Párrafo de contenido...</p>
<h2>Otra subsección</h2>
<p>Más párrafos de contenido...</p>
</body>
</html>
```

Los navegadores muestran el ejemplo anterior de la siguiente manera:



Los navegadores asignan de forma automáticamente el tamaño del título de cada sección en función de su importancia. Así, los títulos de sección `<h1>` se muestran con el tamaño de letra más grande, ya que son el nivel jerárquico superior, mientras que los títulos de sección `<h6>` se visualizan con un tamaño de letra muy pequeño, adecuado para el nivel jerárquico de menor importancia.

Evidentemente, el aspecto que los navegadores aplican por defecto a los títulos de sección se puede modificar utilizando las hojas de estilos de CSS. La siguiente imagen muestra el tamaño por defecto con el que los navegadores muestran cada titular:



2.6 Estructura y disposición

La forma más sencilla de estructurar un texto consiste en separarlo por párrafos. Además, HTML permite incluir títulos que delimitan cada una de las secciones.

Párrafos

Una de las etiquetas más utilizadas de HTML es la etiqueta `<p>`, que permite definir los párrafos que forman el texto de una página. Para delimitar el texto de un párrafo, se encierra ese texto con la etiqueta `<p>`, como muestra el siguiente ejemplo:

```
<html>
```

```
<head>
```

```
<title>Ejemplo de texto estructurado con párrafos</title>
```

```
</head>
```

```
<body>
```

```
<p>Este es el texto que forma el primer párrafo de la página.
```

Los párrafos pueden ocupar varias líneas y el navegador se encarga de ajustar su longitud al tamaño de la ventana.</p>

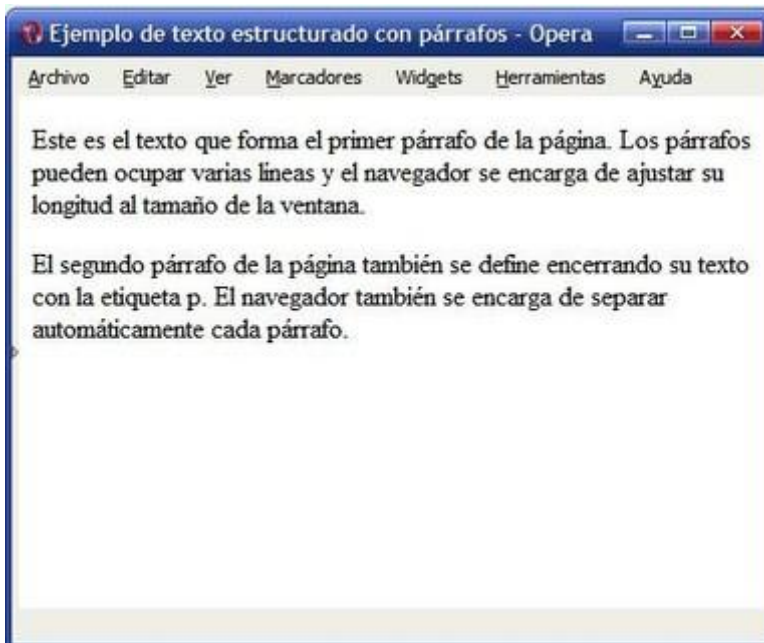
```
<p>El segundo párrafo de la página también se define encerrando su texto con la etiqueta p. El navegador también se encarga de
```

```
se parar automáticamente cada párrafo.</p>
```

```
</body>
```

```
</html>
```

El ejemplo anterior se visualiza de la siguiente manera en cualquier navegador:



La siguiente tabla recoge la definición formal de la etiqueta <p>:

Atributos comunes	básicos, i18n y eventos
Atributos propios	-
Tipo de elemento	Bloque
Descripción	Delimita el contenido de un párrafo de texto

Los párrafos creados con HTML son elementos de bloque, por lo que siempre ocupan toda la anchura de la ventana del navegador. Además, no tienen atributos específicos, pero sí que se les pueden asignar los atributos comunes de HTML básicos, de internacionalización y de eventos.

Secciones

Las páginas HTML habituales suelen tener una estructura más compleja que la que se puede crear solamente mediante párrafos. De hecho, es habitual que las páginas se dividan en diferentes secciones jerárquicas.

Los títulos de sección se utilizan para delimitar el comienzo de cada sección de la página. HTML permite crear secciones de hasta seis niveles de importancia. De esta forma, aunque una página puede definir cualquier número de secciones, sólo puede incluir seis niveles jerárquicos.

Las etiquetas que definen los títulos de sección son <h1>, <h2>, <h3>, <h4>, <h5> y <h6>. La etiqueta <h1> es la de mayor importancia y por tanto se utiliza para definir los titulares de la página. La importancia del resto de etiquetas es descendiente, de forma que la etiqueta <h6> es la que se utiliza para delimitar las secciones menos importantes de la página.

A continuación se muestra la definición formal de la etiqueta <h1>, siendo idéntica la definición del resto de etiquetas referidas a los títulos de sección:

Etiqueta	<h1>
Atributos comunes	básicos, i18n y eventos
Atributos propios	-
Tipo de elemento	Bloque
Descripción	Define los títulos de las secciones de mayor importancia de la página.

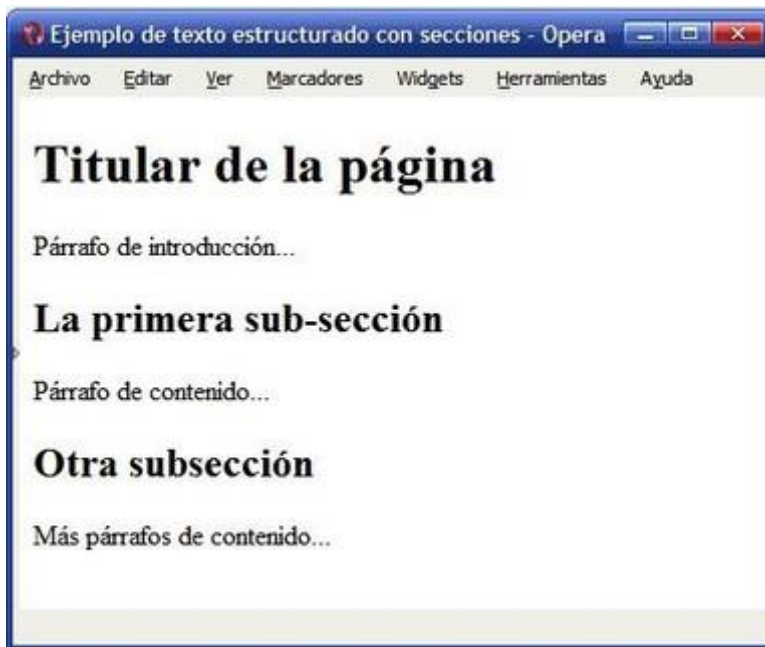
Al igual que la etiqueta <p>, las etiquetas de título de sección son elementos de bloque y no tienen atributos específicos.

Las etiquetas <h1>, ..., <h6> definen títulos de sección, no secciones completas. Por este motivo, no es necesario encerrar los contenidos de una sección con su etiqueta correspondiente. Solamente se debe encerrar con las etiquetas <h1>, ..., <h6> los títulos de cada sección.

El siguiente ejemplo muestra el uso de las etiquetas de título de sección:

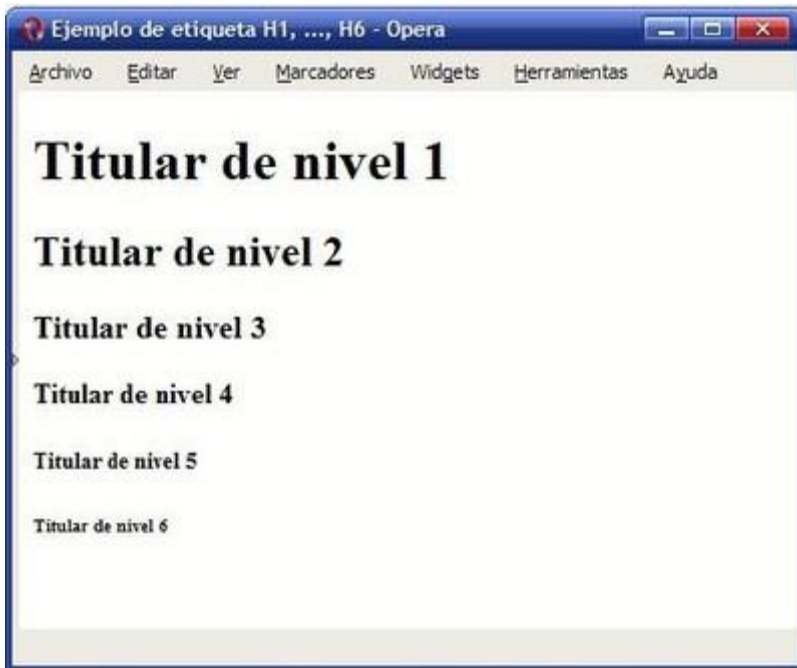
```
<html>
<head>
<title>Ejemplo de texto estructurado con secciones</title>
</head>
<body>
<h1>Titular de la página</h1>
<p>Párrafo de introducción...</p>
<h2>La primera sub-sección</h2>
<p>Párrafo de contenido...</p>
<h2>Otra subsección</h2>
<p>Más párrafos de contenido...</p>
</body>
</html>
```

Los navegadores muestran el ejemplo anterior de la siguiente manera:



Los navegadores asignan de forma automáticamente el tamaño del título de cada sección en función de su importancia. Así, los títulos de sección <h1> se muestran con el tamaño de letra más grande, ya que son el nivel jerárquico superior, mientras que los títulos de sección <h6> se visualizan con un tamaño de letra muy pequeño, adecuado para el nivel jerárquico de menor importancia.

Evidentemente, el aspecto que los navegadores aplican por defecto a los títulos de sección se puede modificar utilizando las hojas de estilos de CSS. La siguiente imagen muestra el tamaño por defecto con el que los navegadores muestran cada titular:



2.7 Formularios

Los formularios más sencillos se pueden crear utilizando solamente dos etiquetas: `<form>` y `<input>`. Si se considera el formulario que muestra la siguiente imagen:



Figura 8.1 Formulario sencillo definido con las etiquetas `form` e `input`

El código HTML necesario para definir el formulario anterior se muestra a continuación:

```
<html>
```

```
<head><title>Ejemplo de formulario sencillo</title></head>
```

```
<body>
```

```
<h3>Formulario muy sencillo</h3>
```

```
<form action="http://www.librosweb.es/maneja_formulario.php" method="post">
```

Escribe tu nombre:

```
<input type="text" name="nombre" value="" />
```

```
<br/>
```

```
<input type="submit" value="Enviar" />
```

```
</form>
```

```
</body>
```

```
</html>
```

La etiqueta `<form>` encierra todos los contenidos del formulario (botones, cuadros de texto, listas desplegables) y la etiqueta `<input>` permite definir varios tipos diferentes de elementos (botones y cuadros de texto).