

## Tema 3.

# Texto

La mayor parte del contenido de las páginas HTML habituales está formado por texto. Por este motivo, es muy importante conocer los elementos y etiquetas que define HTML para el manejo del texto.

HTML define etiquetas para **estructurar** el contenido en secciones y párrafos y define otras etiquetas para **marcar** elementos importantes dentro del texto.

La tarea inicial del editor de contenidos HTML consiste en estructurar el texto original definiendo sus párrafos, titulares y títulos de sección, como se muestra en la siguiente imagen:



**Figura 3.1** Resultado de estructurar un texto sencillo

Una vez definida la estructura básica de los contenidos de la página, el siguiente paso consiste en marcar los diferentes elementos dentro del propio texto: definiciones, abreviaturas, textos importantes, textos modificados, citas a otras referencias, etc.



## 3.1. Estructurar

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.

### 3.1.1. 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. Ejemplo:

```
<html>

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

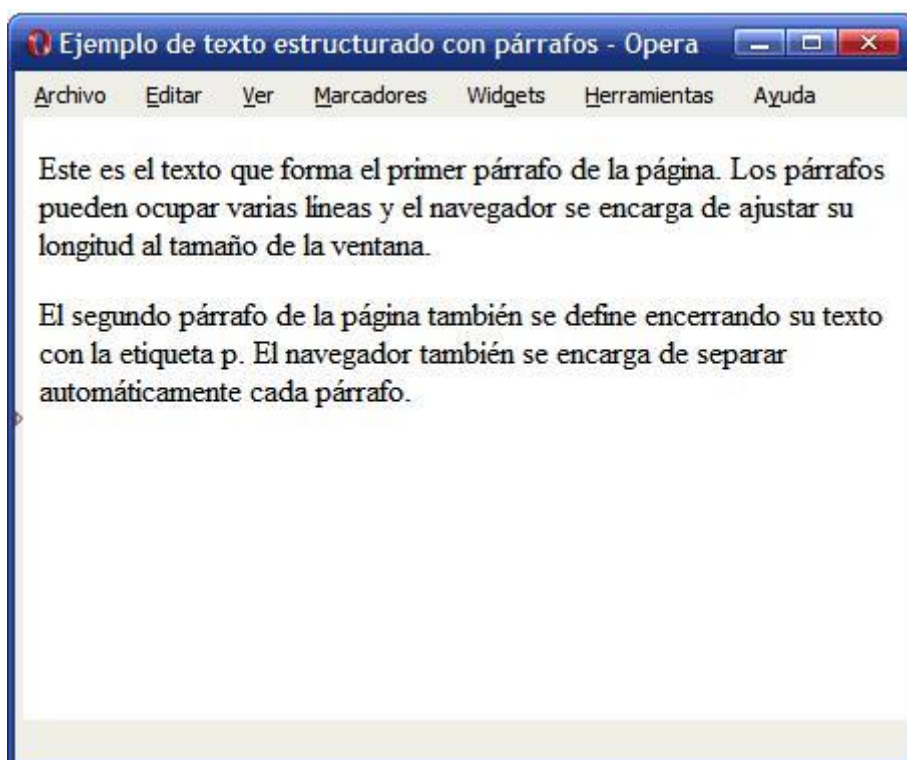
</head>

<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:



**Figura 3.3** Ejemplo de texto HTML estructurado con párrafos

### 3.1.2. Secciones

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.

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.

# Encabezado H1

## Encabezado H2

### Encabezado H3

#### Encabezado H4

##### Encabezado H5

###### Encabezado H6

Los títulos de sección `<h1>` se muestran con el tamaño de letra más grande, mientras que los títulos de sección `<h6>` se visualizan con un tamaño de letra muy pequeño.

Evidentemente, el aspecto que los navegadores aplican por defecto a los títulos de sección se puede modificar utilizando las hojas de estilos de CSS.

## 3.2. Marcado básico de texto

Hace referencia a palabras en negrita o cursiva, anotaciones y correcciones, citas a otros documentos externos, etc. HTML proporciona varias etiquetas para *marcar* cada uno de los diferentes tipos de texto.

Entre las etiquetas más utilizadas para marcar texto se encuentran `<em>` y `<strong>`.

`<em>` `<i>` → Realza la importancia del texto que encierra (cursiva)

`<strong>` `<b>` → Realza con la máxima importancia el texto que encierra (negrita)

La etiqueta `<em>` marca un texto indicando que su importancia es mayor que la del resto del texto. La etiqueta `<strong>` indica que un determinado texto es de la mayor importancia dentro de la página. Ejemplo:

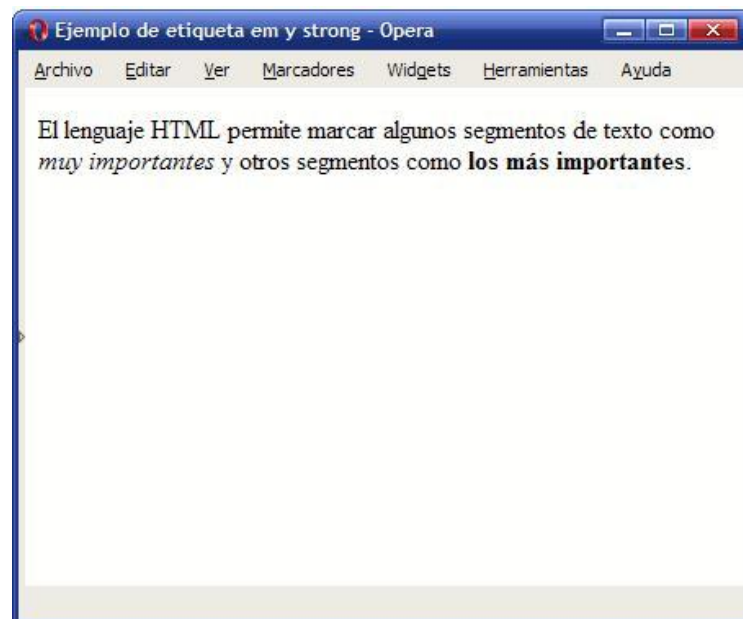
```
<html>

<head>
<title>Ejemplo de etiqueta em y
strong</title> </head>

<body>
<p>El lenguaje HTML permite marcar algunos segmentos de
texto como <em>muy importantes</em> y otros segmentos
como <strong>los más importantes</strong>.</p>
</body>

</html>
```

Por defecto, los navegadores muestran los elementos `<em>` en cursiva para hacer evidente su importancia y muestran los elementos `<strong>` en negrita, para indicar que son los más importantes:



`<ins>` `<u>` → Pone un subrayado al texto

`<del>` `<s>` → Tacha el texto mediante una línea a media altura

Se utilizan para indicar que hemos eliminado una parte del texto original (tachado) para sustituirlo por otro (subrayado)

Ejemplo:

```

<html>
<head><title>Ejemplo de etiqueta ins y
del</title></head> <body>

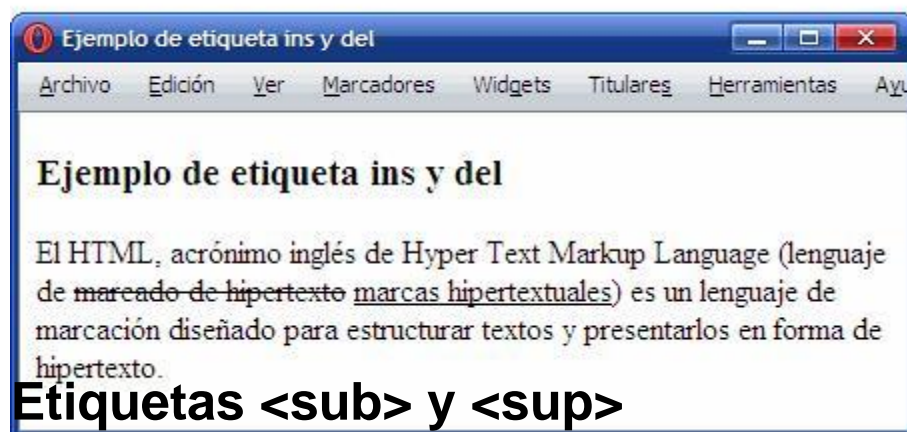
<h3>Ejemplo de etiqueta ins y del</h3>

<p>El HTML, acrónimo inglés de Hyper Text Markup Language (lenguaje de
<del>marcado de hipertexto</del> <ins> marcas hipertextuales</ins>) es un
lenguaje de marcación diseñado para estructurar textos y presentarlos en
forma de hipertexto.</p>

</body>
</html>

```

Los navegadores muestran el ejemplo anterior de la siguiente manera:



Se utilizan para poner subíndices y superíndices

Por otra parte, en muchos tipos de páginas (artículos, noticias) es habitual citar literalmente un texto externo. HTML define la etiqueta `<blockquote>` para incluir citas textuales en las páginas web.

**<blockquote>** → Se emplea para indicar que el texto que encierra es una cita textual de otro texto externo

Ejemplo:

```

<html>
<head><title>Ejemplo de etiqueta blockquote</title></head>

<body>
<p>Según el W3C, el valor del
atributo <em>cite</em> en las etiquetas
<strong>blockquote</strong> tiene el siguiente significado:</p>

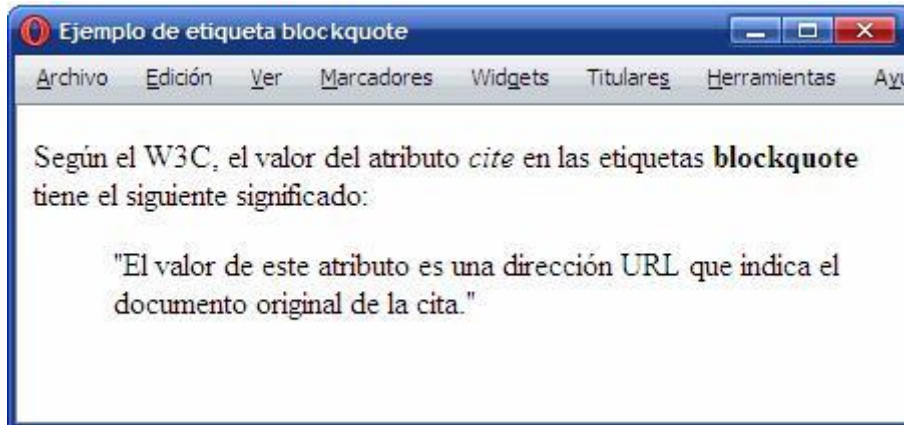
```

```
<blockquote cite="http://www.w3.org/TR/html401/struct/text.html">"El
valor de este atributo
es una dirección URL que indica el documento original de la
cita."</blockquote> </body>

</html>
```

---

El aspecto que muestra el ejemplo anterior en cualquier navegador es el siguiente:



**Figura 3.9** Ejemplo de uso de la etiqueta `blockquote`

Para indicar de forma clara que el texto es una cita externa, los navegadores muestran por defecto el texto del elemento `<blockquote>` con un gran margen en la parte izquierda.

### 3.3. Marcado avanzado de texto

**<abbr>** → Se emplea para marcar las abreviaturas del texto y proporcionar el significado de esas abreviaturas

**<acronym>** → Se emplea para marcar las siglas o acrónimos del texto y proporcionar el significado de esas siglas

En ambos casos, el atributo `title` se puede utilizar para incluir el significado completo de la abreviatura o sigla.

Ejemplo:

---

```
<html>

<head>
<title>Ejemplo de etiqueta
acronym</title> </head>
```

```

<body>
<p>El lenguaje <acronym title="HyperText Markup
Language">HTML</acronym> es estandarizado
por el <acronym title="World Wide Web
Consortium">W3C</acronym>.</p> </body>

</html>

```

La mayoría de navegadores muestran por defecto un borde inferior punteado para todos los elementos `<abbr>` y `<acronym>`. Al posicionar el puntero del ratón sobre la palabra subrayada, el navegador muestra un pequeño recuadro (llamado *tooltip* en inglés) con el valor del atributo `title`:



**<cite>** → Se emplea para marcar una cita o una referencia a otras fuentes. Se pone en cursiva

En ocasiones, no está clara la diferencia entre `<cite>` y `<blockquote>`. El elemento `<cite>` marca el autor de la cita (persona, documento, etc.) y `<blockquote>` marca el contenido de la propia cita. En el siguiente ejemplo, `<blockquote>` encierra el contenido de una frase célebre y `<cite>` encierra el nombre de su autor:

```

Como dijo <cite>Mahatma Gandhi</cite>:
<blockquote>Vive como si fueras a morir mañana y aprende como si
fueras a vivir para siempre.</blockquote>

```



## 3.4. Marcado genérico de texto

A veces no disponemos de una etiqueta concreta para marcar un determinado contenido de texto. Si se considera el siguiente ejemplo:

---

```
Importante: si quiere ponerse en contacto con la empresa ACME, puede hacerlo en el
teléfono 900 555 555 o a través de la dirección de correo electrónico contacto@acme.org
```

El texto del ejemplo anterior contiene elementos de texto importantes, siglas, números de teléfono y direcciones de correo electrónico. XHTML define la etiqueta `<strong>` para marcar los elementos importantes y `<acronym>` para marcar las siglas:

---

```
<strong>Importante</strong>: si quiere ponerse en contacto con la
empresa <acronym>ACME</acronym>, puede hacerlo en el teléfono 900 555
555 o a través de la dirección de correo electrónico contacto@acme.org
```

Desafortunadamente, XHTML no define ninguna etiqueta específica para marcar números de teléfono o direcciones de correo electrónico. De la misma forma, no define etiquetas para otros posibles elementos que se pueden encontrar en los contenidos de texto.

Por este motivo, el estándar HTML/XHTML incluye una etiqueta llamada `<span>` que se emplea para marcar cualquier elemento que no se puede marcar con las otras etiquetas definidas.

Siguiendo con el ejemplo anterior, la etiqueta `<span>` se utiliza para marcar el teléfono y la dirección de correo electrónico:

```
<strong>Importante</strong>: si quiere ponerse en contacto con la empresa
<acronym>ACME</acronym>, puede hacerlo en el teléfono <span>900 555 555</span> o
a través de la dirección de correo electrónico <span>contacto@acme.org</span>
```

La etiqueta `<span>` se visualiza por defecto con el mismo aspecto que el texto normal. Por tanto es habitual utilizar esta etiqueta junto con los atributos `id` y `class` para modificar posteriormente su aspecto con CSS:

---

```
<strong>Importante</strong>: si quiere ponerse en contacto con la
empresa <acronym>ACME</acronym>, puede hacerlo en el teléfono <span
class="telefono">900 555 555</span> o a través de la dirección de
correo electrónico <span class="email">contacto@acme.org</span>
```

## 3.5. Espacios en blanco y nuevas líneas

HTML considera *espacio en blanco* a los espacios en blanco, los tabuladores, los retornos de carro y el carácter de nueva línea (**ENTER** o **Intro**).

El siguiente ejemplo ilustra este comportamiento:

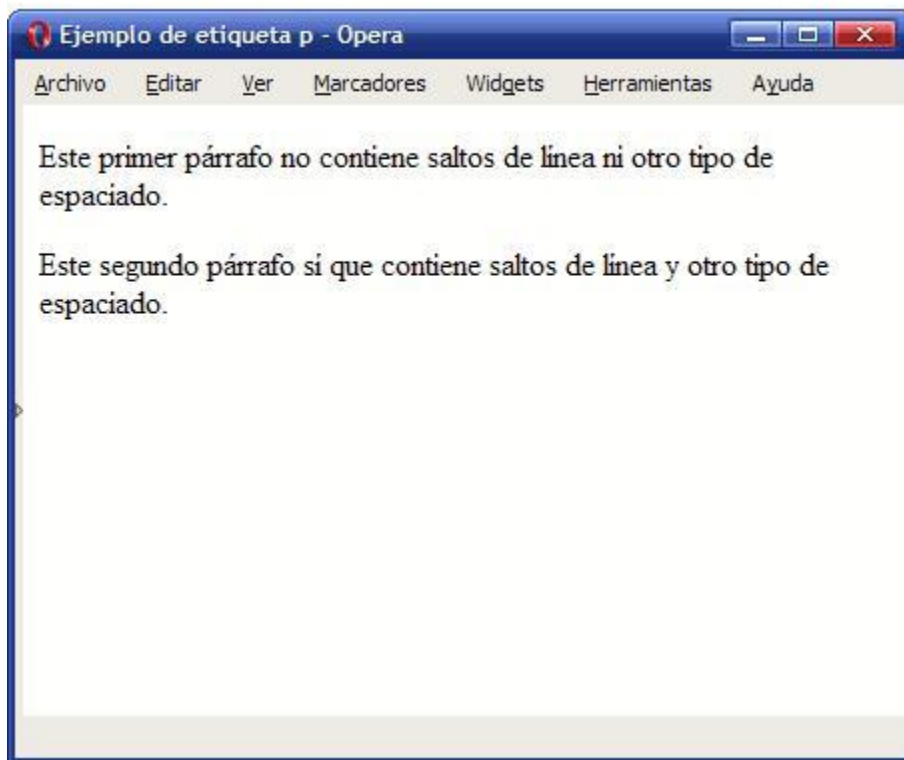
---

```
<html>
<head><title>Ejemplo de etiqueta
p</title></head> <body>
<p>Este primer párrafo no contiene saltos de línea ni otro tipo de espaciado.</p>
```



```
<p>Este segundo párrafo sí que  
contiene saltos de  
línea  
y otro tipo de espaciado.</p>  
</body>  
</html>
```

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



Los dos párrafos de la imagen anterior se ven idénticos, aunque el segundo párrafo incluye varios espacios en blanco y está escrito en varias líneas diferentes. La razón de este comportamiento es que HTML ignora todos los espacios en blanco *sobrantes*, es decir, todos los espacios en blanco que no son el espacio en blanco que separa las palabras.

No obstante, HTML proporciona varias alternativas para poder incluir tantos espacios en blanco y tantas nuevas líneas como sean necesarias dentro del contenido textual de las páginas.

### 3.5.1. Nuevas líneas

**<br>** → Fuerza al navegador a insertar una nueva línea

La etiqueta `<br>` es que es una etiqueta vacía, es decir, no encierra ningún texto. De esta forma, la etiqueta debe abrirse y cerrarse de forma consecutiva: `<br></br>`.

En estos casos, HTML permite utilizar un atajo para indicar que una etiqueta se está abriendo y cerrando de forma consecutiva: `<br />` (también se puede escribir como `<br />`).

Utilizando la etiqueta `<br>` se puede rehacer el ejemplo anterior para que respete las líneas que forman el segundo párrafo:

```
<html>

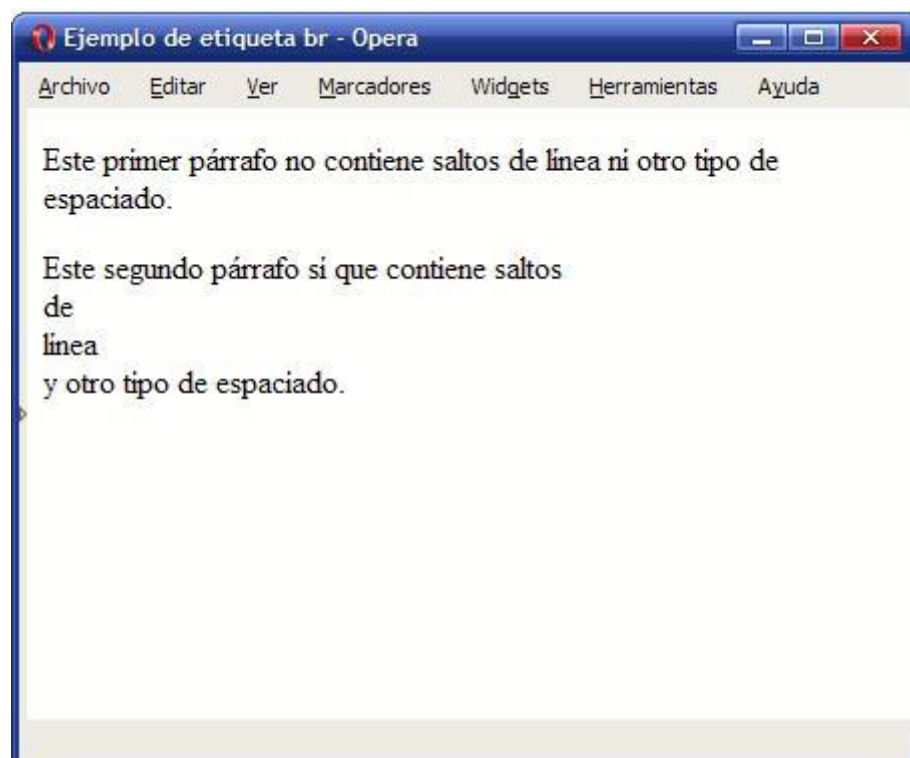
<head>
<title>Ejemplo de etiqueta
br</title> </head>

<body>
<p>Este primer párrafo no contiene saltos de línea ni otro tipo de espaciado.</p>

<p>Este segundo párrafo sí que contiene
saltos <br/> de <br/>
línea <br/>
y otro tipo de espaciado.</p>
</body>

</html>
```

El navegador ahora sí que muestra correctamente las nuevas líneas que se querían insertar:



Esta etiqueta solo se utilizara cuando sea imprescindible insertar un salto de página

### 3.5.2. Espacios en blanco

Para incluir espacios en blanco adicionales, se debe sustituir cada nuevo espacio en blanco por el texto `&nbsp;` (es importante incluir el símbolo `&` al principio y el símbolo `;` al final).

Así, el código HTML del ejemplo anterior se debe rehacer para incluir los espacios en blanco adicionales:

```
<html>

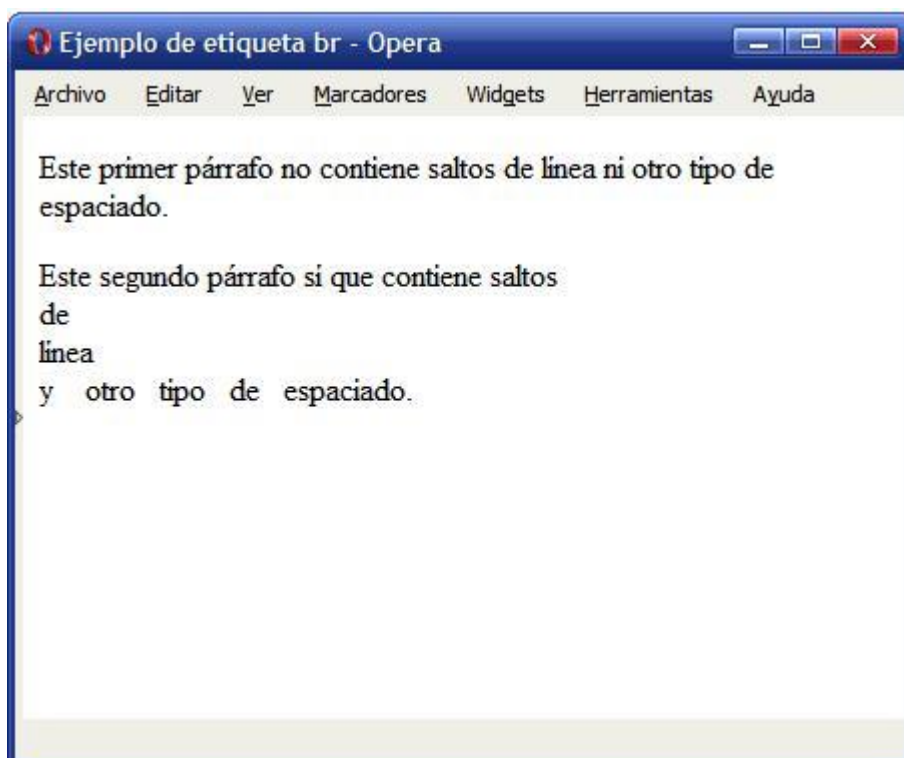
<head>
<title>Ejemplo de entidad
&nbsp;</title> </head>

<body>
<p>Este primer párrafo no contiene saltos de línea ni otro tipo de espaciado.</p>

<p>Este segundo párrafo sí que contiene
saltos <br/> de <br/>
línea <br/>
y &nbsp;&nbsp;&nbsp; otro &nbsp;&nbsp;&nbsp; tipo &nbsp;&nbsp;&nbsp; de &nbsp;&nbsp;&nbsp;
espaciado.</p> </body>

</html>
```

Ahora el navegador sí que muestra correctamente los espacios en blanco (y las nuevas líneas) del segundo párrafo:



Cada texto `&nbsp;` solamente equivale a un espacio en blanco, por lo que se deben escribir tantos `&nbsp;` seguidos como espacios en blanco seguidos existan en el texto.

### 3.5.3. Texto preformateado

En ocasiones, es necesario mostrar los espacios en blanco de un texto que no se puede modificar. Se trata de un caso habitual cuando una página web debe mostrar directamente el texto generado por alguna aplicación.

En estos casos, se puede utilizar la etiqueta `<pre>`, que muestra el texto tal y como se ha escrito, respetando todos los espacios en blanco y todas las nuevas líneas.

**`<pre>`** → Muestra el texto que encierra tal y como está escrito (respetando los espacios en blanco)

El siguiente ejemplo muestra el uso de la etiqueta `<pre>`:

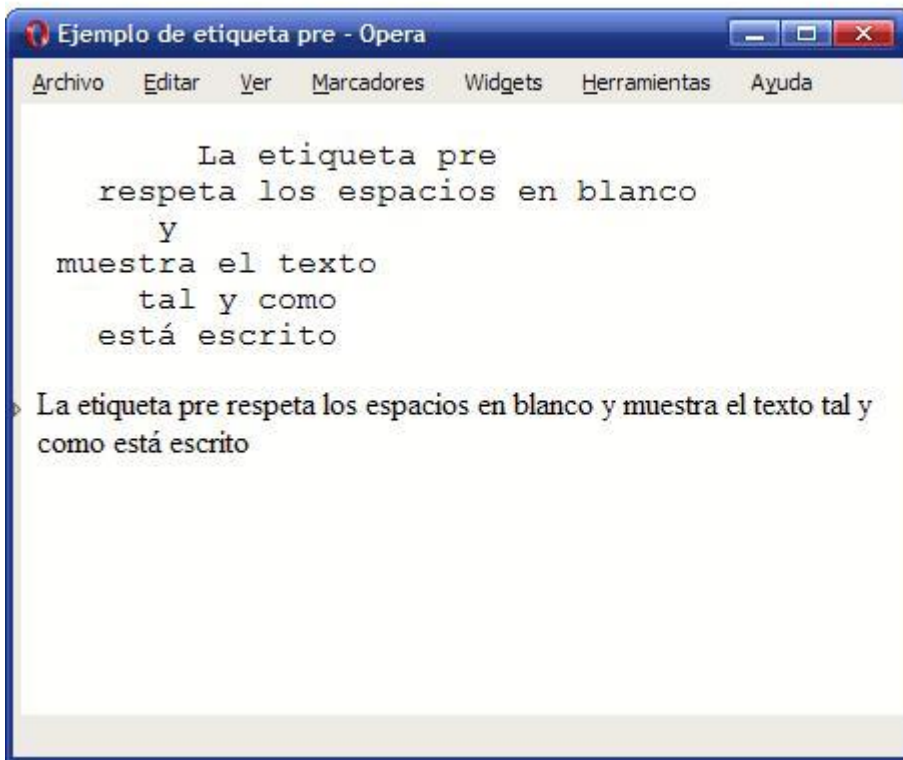
```
<html>
<head><title>Ejemplo de etiqueta pre</title></head>

<body>
<pre>
    La etiqueta pre
    respeta los espacios en
    blanco y
    muestra el texto
    tal y como está
    escrito
</pre>

<p>
    La etiqueta pre
    respeta los espacios en
    blanco y
    muestra el texto
    tal y como está
    escrito
</p>

</body>
</html>
```

El ejemplo anterior incluye el mismo texto (con espacios en blanco y varias líneas) dentro de una etiqueta `<pre>` y dentro de una etiqueta `<p>`. Las diferencias visuales en un navegador son muy evidentes:



**Figura 3.16** Ejemplo de uso de la etiqueta pre

El primer texto se ve en pantalla tal y como se ha escrito, respetando todos los espacios en blanco y todas las nuevas líneas. El segundo texto se ve como un párrafo normal, ya que HTML ha eliminado todos los espacios en blanco sobrantes.

Los elementos `<pre>` son *especiales*, ya que los navegadores les aplican las siguientes reglas:

- Mantienen todos los espacios en blanco (tabuladores, espacios y nuevas líneas)
- Muestra el texto con un tipo de letra especial, denominado "*de ancho fijo*", ya que todas sus letras son de la misma anchura
- No se ajusta la longitud de las líneas (las líneas largas producen un *scroll* horizontal en la ventana del navegador)

## 3.6. Codificación de caracteres

Algunos de los caracteres que se utilizan habitualmente en los textos no se pueden incluir directamente en las páginas web:

- Los caracteres que utiliza HTML para definir sus etiquetas (`<`, `>` y `"`) no se pueden utilizar libremente.
- Los caracteres propios de los idiomas que no son el inglés (`ñ`, `á`, `ç`, `¿`, `¡`, etc.) pueden ser problemáticos dependiendo de la codificación de caracteres utilizada.

La solución consiste en sustituir los caracteres reservados de HTML por unas expresiones llamadas *entidades HTML* y que representan a cada carácter:

Entidad	Carácter	Descripción	Traducción
---------	----------	-------------	------------

<	<	less than	signo de menor que
>	>	more than	signo de mayor que
&	&	ampersand	ampersand
"	"	quotation mark	comillas
	(espacio en blanco)	non-breaking space	espacio en blanco
'	'	apostrophe	apóstrofo

De esta forma, si se considera el siguiente texto:

Los caracteres <, >, " y & pueden dar problemas con los textos en HTML

Para mostrar correctamente el texto anterior en una página HTML, se debe sustituir cada carácter especial por su entidad HTML:

```
<p>Los caracteres &lt;, &gt;, &quot; y &amp; pueden dar problemas con  
los textos en HTML</p>
```

Si el editor, el diseñador, el entorno de desarrollo, el servidor y el navegador utilizan la misma codificación de caracteres, los caracteres propios de los idiomas se pueden escribir directamente:

```
<p>Este párrafo contiene caracteres acentuados y se almacena en formato UTF-8</p>
```

La palabra **párrafo** del ejemplo anterior incluye la letra **á**. Si el editor HTML del diseñador utiliza la codificación UTF-8, el entorno de desarrollo del programador también utiliza UTF-8, el servidor web sirve las páginas con esa codificación y el navegador del usuario es capaz de visualizar las páginas con formato UTF-8, el texto anterior se verá correctamente en el navegador del usuario.

Sin embargo, muchas veces no es posible que todos los procesos involucrados utilicen la misma codificación de caracteres. Por limitaciones técnicas o por decisiones de los diseñadores y programadores, los textos pueden pasar de codificación UTF-8 a codificación ISO-8859 en cualquier momento. Si se produce este cambio sin realizar una conversión correcta, el navegador del usuario mostrará caracteres extraños en todos los acentos y en todas las letras como la **ñ**.

La solución para asegurar que todos estos caracteres potencialmente problemáticos se van a visualizar correctamente en el navegador del usuario consiste en sustituir cada carácter problemático por su entidad HTML:

Entidad	Carácter	Descripción oficial
&ntilde;	ñ	latin letter n with tilde
&Ntilde;	Ñ	latin capital n letter with tilde

<code>&amp;aacute;</code>	á	a acute
<code>&amp;eacute;</code>	é	e acute
<code>&amp;iacute;</code>	í	i acute
<code>&amp;oacute;</code>	ó	o acute
<code>&amp;uacute;</code>	ú	u acute
<code>&amp;Aacute;</code>	Á	A acute
<code>&amp;Eacute;</code>	É	E acute
<code>&amp;Iacute;</code>	Í	I acute
<code>&amp;Oacute;</code>	Ó	O acute
<code>&amp;Uacute;</code>	Ú	U acute
<code>&amp;euro;</code>	€	euro

Así, el párrafo de texto del ejemplo anterior, se podría escribir de la siguiente manera:

```
<p>Este párrafo contiene caracteres acentuados y se almacena en formato UTF-8</p>
```

Si se utilizan las entidades HTML en vez de los caracteres problemáticos, es indiferente pasar de una codificación de caracteres a otra diferente.

## <Address>

Se utiliza para indicar información de contacto. Se suele emplear en el pie de la página.