

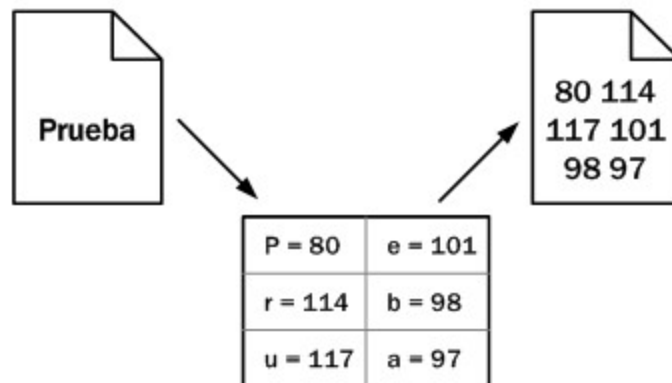
## TEMA 2. CARACTERÍSTICAS BÁSICAS.

1.	LENGUAJES DE ETIQUETAS.....	2
2.	EL PRIMER DOCUMENTO HTML.....	3
3.	ETIQUETAS Y ATRIBUTOS.....	5
4.	ELEMENTOS HTML.....	6
5.	SINTAXIS DE LAS ETIQUETAS.....	7

## 1. LENGUAJES DE ETIQUETAS

Uno de los retos iniciales a los que se tuvo que enfrentar la informática fue el de cómo almacenar la información en los archivos digitales. Como los primeros archivos sólo contenían texto sin formato, la solución utilizada era muy sencilla: se codificaban las letras del alfabeto y se transformaban en números.

De esta forma, para almacenar un contenido de texto en un archivo electrónico, se utiliza una tabla de conversión que transforma cada carácter en un número. Una vez almacenada la secuencia de números, el contenido del archivo se puede recuperar realizando el proceso inverso.



**Figura 2.1** Ejemplo sencillo de codificación de caracteres

El proceso de transformación de caracteres en secuencias de números se denomina **codificación de caracteres** y cada una de las tablas que se han definido para realizar la transformación se conocen con el nombre de **páginas de código**. Una de las codificaciones más conocidas (y una de las primeras que se publicaron) es la codificación ASCII.

Si por ejemplo se quiere dividir el texto en párrafos y se desea dar especial importancia a algunas palabras, se podría indicar de la siguiente manera:

<parrafo>

Contenido de texto con <importante>algunas palabras</importante> resaltadas de forma especial.

</parrafo>

El principio de un párrafo se indica mediante la palabra <parrafo> y el final de un párrafo se indica mediante la palabra </parrafo>. De la misma manera, para asignar más importancia a ciertas palabras del texto, se encierran entre <importante> y </importante>.

El proceso de indicar las diferentes partes que componen la información se denomina **marcar** (*markup* en inglés). Cada una de las palabras que se emplean para marcar el inicio y el final de una sección se denominan **etiquetas**.

Aunque existen algunas excepciones, en general las etiquetas se indican por pares y se forman de la siguiente manera:

- Etiqueta de apertura: carácter `<`, seguido del nombre de la etiqueta (sin espacios en blanco) y terminado con el carácter `>`
- Etiqueta de cierre: carácter `<`, seguido del carácter `/`, seguido del nombre de la etiqueta (sin espacios en blanco) y terminado con el carácter `>`

Así, la estructura típica de las etiquetas HTML es:

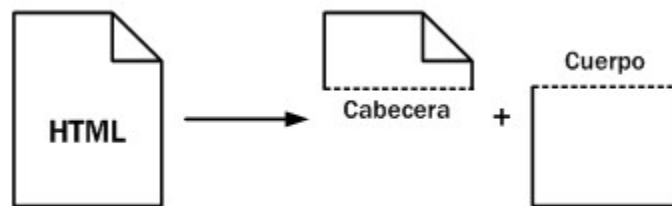
`<nombre_etiqueta> ... </nombre_etiqueta>`

HTML es un **lenguaje de etiquetas** (también llamado **lenguaje de marcado**) y las páginas web habituales están formadas por cientos o miles de pares de etiquetas. De hecho, las letras "ML" de la sigla HTML significan "*markup language*", que es como se denominan en inglés a los *lenguajes de marcado*. Además de HTML, existen muchos otros lenguajes de etiquetas como XML, SGML, DocBook y MathML.

La principal ventaja de los lenguajes de etiquetas es que son muy sencillos de leer y escribir por parte de las personas y de los sistemas electrónicos. La principal desventaja es que pueden aumentar mucho el tamaño del documento, por lo que en general se utilizan etiquetas con nombres muy cortos.

## 2. EL PRIMER DOCUMENTO HTML

Las páginas HTML se dividen en dos partes: la cabecera y el cuerpo. La cabecera incluye información sobre la propia página, como por ejemplo su título y su idioma. El cuerpo de la página incluye todos sus contenidos, como párrafos de texto e imágenes.



**Figura 2.2** Esquema de las partes que forman un documento HTML

El cuerpo (llamado *body* en inglés) contiene todo lo que el usuario ve en su pantalla y la cabecera (llamada *head* en inglés) contiene todo lo que no se ve (con la única excepción del título de la página, que los navegadores muestran como título de sus ventanas).

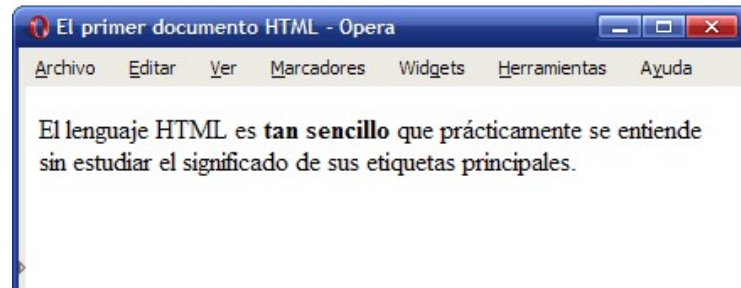
A continuación, se muestra el código HTML de una página web muy sencilla:

```
<html>
  <head>
    <title>El primer documento HTML</title>
  </head>
  <body>
    <p>El lenguaje HTML es <strong>tan sencillo</strong>
que
    prácticamente se entiende sin estudiar el significado
    de sus etiquetas principales. </p>
  </body>
</html>
```

Si quieres probar este primer ejemplo, debes hacer lo siguiente:

1. Abre un editor de archivos de texto y crea un archivo nuevo
2. Copia el código HTML mostrado anteriormente y pégalo tal cual en el archivo que has creado
3. Guarda el archivo con el nombre que quieras, pero con la extensión .html

Después de crear el archivo con el contenido HTML, ya se puede abrir con cualquier navegador para que se muestre con el siguiente aspecto:

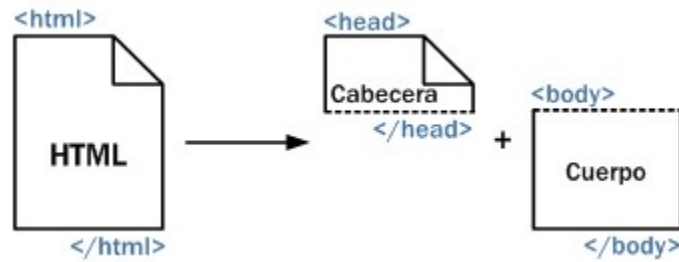


**Figura 2.3** Aspecto que muestra el primer documento HTML en cualquier navegador

Si ya estás viendo tu primera página HTML en el navegador, prueba a pulsar sobre el menú **Ver > Código fuente** y podrás ver el código HTML de la página que está cargada en el navegador. De hecho, puedes ver el código HTML de cualquier página de Internet mediante la opción **Ver > Código fuente**. Prueba a ver el código HTML de tu página preferida y verás cuantas etiquetas puede llegar a tener una página compleja.

Volviendo al código HTML del primer ejemplo, es importante conocer las tres etiquetas principales de un documento HTML (`<html>`, `<head>`, `<body>`):

- `<html>`: indica el comienzo y el final de un documento HTML. Ninguna etiqueta o contenido puede colocarse antes o después de la etiqueta `<html>` (con una sola excepción que se verá más adelante). En el interior de la etiqueta `<html>` se definen la cabecera y el cuerpo del documento HTML y todo lo que se coloque fuera de la etiqueta `<html>` se ignora.
- `<head>`: delimita la parte de la cabecera del documento. La cabecera contiene información sobre el propio documento HTML, como por ejemplo su título y el idioma de la página. Los contenidos indicados en la cabecera no son visibles para el usuario, con la excepción de la etiqueta `<title>`, que se utiliza para indicar el título del documento y que los navegadores lo visualizan en la parte superior izquierda de la ventana del navegador (si no te has fijado anteriormente, vuelve a abrir el primer ejemplo en cualquier navegador y observa dónde se muestra el título de la página).
- `<body>`: delimita el cuerpo del documento HTML. El cuerpo encierra todos los contenidos que se muestran al usuario (párrafos de texto, imágenes, tablas). En general, el `<body>` de un documento contiene cientos de etiquetas HTML, mientras que el `<head>` no contiene más que unas pocas.



**Figura 2.4** Esquema de las etiquetas principales que contiene un documento HTML

### 3. ETIQUETAS Y ATRIBUTOS

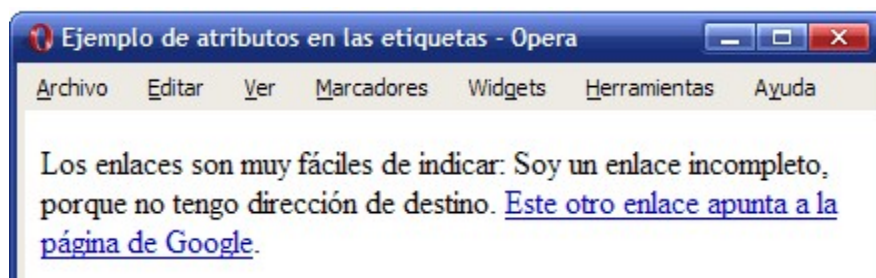
Las etiquetas en HTML nos permiten definir los elementos que componen el documento.

Algunos elementos (etiquetas) como las imágenes y los enlaces requieren cierta información adicional para estar completamente definidos.

La etiqueta `<a>` por ejemplo se emplea para incluir un enlace en una página. Utilizando sólo la etiqueta `<a>` no es posible establecer la dirección a la que apunta cada enlace. Como no es viable crear una etiqueta por cada enlace diferente, la solución consiste en personalizar las etiquetas HTML mediante cierta información adicional llamada **atributos**.

De esta forma, se utiliza la misma etiqueta `<a>` para todos los enlaces de la página y se utilizan los atributos para indicar la dirección a la que apunta cada enlace.

```
<html>
  <head>
    <title>Ejemplo de atributos en las etiquetas</title>
  </head>
  <body>
    <p>
      <a>Soy un enlace incompleto, porque no tengo dirección de
      destino</a>.
      <a href="http://www.google.com">
        Este otro enlace apunta a la página de Google
      </a>
    </p>
  </body>
</html>
```



**Figura 2.6** Los atributos permiten personalizar las etiquetas HTML

El primer enlace del ejemplo anterior no está completamente definido, ya que no apunta a ninguna dirección. El segundo enlace, utiliza la misma etiqueta `<a>`, pero

añade información adicional mediante un atributo llamado `href`. Los atributos se incluyen dentro de la etiqueta de apertura. Por ahora no es importante comprender la etiqueta `<a>` ni el atributo `href`, ya que se explicarán con todo detalle más adelante.

No todos los atributos se pueden utilizar en todas las etiquetas. Por ello, cada etiqueta define su propia lista de atributos disponibles. Además, cada atributo también indica el tipo de valor que se le puede asignar. Si el valor de un atributo no es válido, el navegador ignora ese atributo.

Los atributos pueden escribirse en el orden que quiera el programador, no afecta al resultado.

**Es muy importante no utilizar atributos para dar formato a las etiquetas, esta función es la que realiza CSS.**

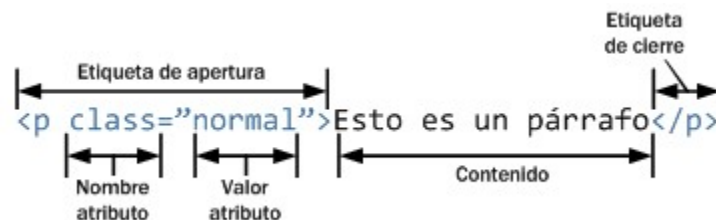
#### 4. ELEMENTOS HTML

Además de etiquetas y atributos, HTML define el término **elemento** para referirse a las partes que componen los documentos HTML.

Aunque en ocasiones se habla de forma indistinta de "elementos" y "etiquetas", en realidad un elemento HTML es mucho más que una etiqueta, ya que está formado por:

- Una etiqueta de apertura.
- Cero o más atributos.
- Texto encerrado por la etiqueta.
- Una etiqueta de cierre.

El texto encerrado por la etiqueta es opcional, ya que algunas etiquetas de HTML no pueden encerrar ningún texto. El siguiente esquema muestra un elemento HTML, formado por una etiqueta `<p>`, atributos y contenidos de texto:



**Figura 2.7** Esquema de las partes que componen un elemento HTML

La estructura mostrada en el esquema anterior es un elemento HTML ya que comienza con una etiqueta de apertura (`<p>`), contiene cero o más atributos (`class="normal"`), dispone de un contenido de texto (Esto es un párrafo) y finaliza con una etiqueta de cierre (`</p>`).

Por tanto, si una página web tiene dos párrafos de texto, la página contiene dos elementos y cuatro etiquetas (dos etiquetas `<p>` de apertura y dos etiquetas `</p>` de cierre). De todas formas, aunque estrictamente no son lo mismo, es habitual intercambiar las palabras "elemento" y "etiqueta".

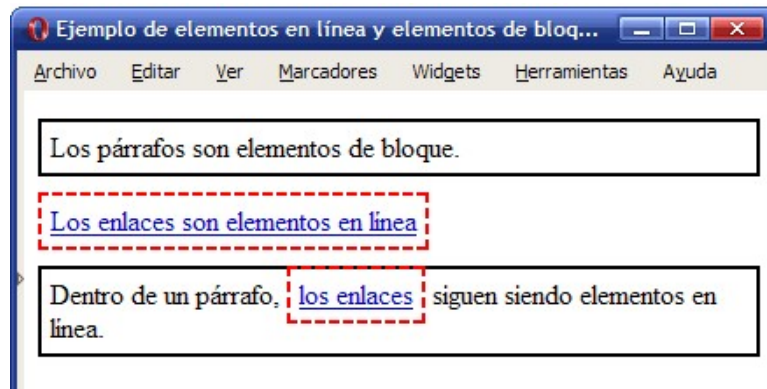
Por otra parte, el lenguaje HTML clasifica a todos los elementos en dos grupos: elementos **en línea** (*inline elements* en inglés) y elementos de **bloque** (*block elements* en inglés).

La principal diferencia entre los dos tipos de elementos es la forma en la que ocupan el espacio disponible en la página. Los elementos de bloque siempre empiezan en una nueva línea y ocupan todo el espacio disponible hasta el final de la línea, aunque sus contenidos no lleguen hasta el final de la línea. Por su parte, los elementos en línea sólo ocupan el espacio necesario para mostrar sus contenidos.

Si se considera el siguiente ejemplo:

```
<html>
  <head>
    <title>Ejemplo de elementos en línea y elementos de bloque</title>
  </head>
  <body>
    <p>Los párrafos son elementos de bloque. </p>
    <a href="http://www.google.com">Los enlaces son elementos en
línea</a>
    <p>
      Dentro de un párrafo,
      <a href="http://www.google.com">los enlaces</a>
      siguen siendo elementos en línea.
    </p>
  </body>
</html>
```

La siguiente imagen muestra cómo visualizan los navegadores el código HTML anterior (mediante CSS se han añadido bordes que muestran el espacio ocupado por cada elemento):



**Figura 2.8** Diferencias entre elementos en línea y elementos de bloque

El primer párrafo contiene un texto corto que sólo ocupa la mitad de la anchura de la ventana del navegador. No obstante, el espacio reservado por el navegador para el primer párrafo llega hasta el final de esa línea, por lo que resulta evidente que los elementos `<p>` son elementos de bloque.

Por otra parte, el primer enlace del ejemplo anterior también tiene un texto corto que ocupa solamente la mitad de la anchura de la ventana del navegador. En este caso, el navegador sólo reserva para el enlace el sitio necesario para mostrar sus contenidos. Si se añade otro enlace en esa misma línea, se mostraría a continuación del primer enlace. Por tanto, los elementos `<a>` son elementos en línea.

Por último, el segundo párrafo sigue ocupando todo el espacio disponible hasta el final de cada línea (por ser un elemento de bloque) y el enlace que se encuentra dentro del párrafo sólo ocupa el sitio necesario para mostrar sus contenidos (por ser un elemento en línea).

## 5. SINTAXIS DE LAS ETIQUETAS.

El lenguaje HTML original era muy permisivo en su sintaxis, por lo que era posible escribir sus etiquetas y atributos de muchas formas diferentes. Las etiquetas por ejemplo podían escribirse en mayúsculas, en minúsculas e incluso combinando mayúsculas y minúsculas. El valor de los atributos de las etiquetas se podía indicar con y sin comillas ("). Además, el orden en el que se abrían y cerraban las etiquetas no era importante.

La flexibilidad de HTML puede parecer un aspecto positivo, pero el resultado final son páginas con un código HTML desordenado, difícil de mantener y muy poco profesional.

A continuación, se muestran algunas restricciones básicas

**Las etiquetas se tienen que cerrar de acuerdo a como se abren:**

### CORRECTO

```
<p>Este es un párrafo con <a>un enlace</a></p>
```

### INCORRECTO

```
<p>Este es un párrafo con <a>un enlace</p></a>
```

**Los nombres de las etiquetas y atributos siempre se escriben en minúsculas:**

### CORRECTO

```
<p>Este es un párrafo con <a href="http://www.google.com">un  
enlace</a></p>
```

### INCORRECTO

```
<P>Este es un párrafo con <A HREF="http://www.google.com">un  
enlace</A></P>
```

**El valor de los atributos siempre se encierra con comillas:**

### CORRECTO

```
<p>Este es un párrafo con <a href="http://www.google.com">un  
enlace</a></p>
```

### INCORRECTO

```
<p>Este es un párrafo con <a href=http://www.google.com>un  
enlace</a></p>
```