

2.4.3. <meta>

Definición: indica un conjunto de propiedades generales del documento como el autor, descripción, palabras clave, herramienta utilizada, tipo de contenido, etc.

Estas propiedades evolucionan continuamente, es decir, se crean nuevas propiedades para ampliar la información asociada al documento, las redes sociales son un ejemplo.

Normalmente estas propiedades pasan desapercibidas para el usuario habitual.

Aparición: sin etiqueta de cierre (V).

Atributos: **lang**, **dir**, **name**, **content**, **http-equiv**, **scheme**.

- **name**: indica el nombre de una propiedad, no hay una lista oficial de propiedades pero podemos citar los siguientes valores: abstract, author, copyright, date, description, distribution, expires, generator, google-site-verification, keywords, language, no-email-collection, organization, rating, reply-to, revisit-after, robots y la familia de etiquetas Dublin Core.
- **content**: contiene el valor de la propiedad indicada antes por **name**, por tanto, estos dos atributos funcionan simultáneamente proporcionando parejas de valores.

Ejemplo:

```
<meta name="description" content="Lenguajes de Marcas">
```

- **http-equiv**: indica una propiedad al navegador en forma de cabecera http, como si el propio servidor http (Web) hubiera generado dicha cabecera, de ahí viene el nombre "http-equivalent".

Valores: cache-control, content-type, set-cookie, content-disposition, pics-label, pragma, refresh, resource-type, content-script-type, content-style-type, window-target.

Tiene muchas utilidades como por ejemplo, indicar el conjunto de caracteres a utilizar en nuestro sitio Web, sin necesidad de modificar el servidor. También funciona en coordinación con el atributo **content** para proporcionar parejas de valores.

Ejemplo:

```
<meta http-equiv="content-type"
  content="text/html; charset=iso-8859-15">
```

Esto indica que el contenido de la página es texto plano HTML con el conjunto de caracteres iso-8859-15, que corresponde al alfabeto latino nº 9.

- **scheme**: indica al navegador que tiene que interpretar los metadatos atendiendo al perfil especificado, para evitar ambigüedad.

Ejemplo:

```
<meta scheme="Europe" name="date" content="24-10-2012">
```

Esto indica que la fecha se lee según el perfil "Europe", es decir: "dd-mm-aaaa".

Ahora vamos a estudiar las principales meta etiquetas en función de su utilidad.

Se clasifican en dos grupos (por sus dos atributos principales, aunque hay más):

- M
E
T
A
- 1) Las **<meta name>**: se utilizan principalmente para optimizar el resultado de los motores de búsqueda, es decir, para ayudar a los buscadores de páginas Web.
La información que manejan no se presenta en pantalla.
- 2) Las **<meta http-equiv>**: se utilizan para dar instrucciones a los navegadores, es decir, ejercen cierto control sobre ellos.
Tienen influencia sobre lo que aparece en pantalla.

1) Meta name

- Información general sobre la página:
`abstract, author, copyright, date, generator`
- Información específica para buscadores:
`description, distribution, keywords, locality, rating, revisit-after, robots`
- Para administración del sitio:
`google-site-verification, reply-to`

2) Meta http-equiv

- Tipo de Contenido:
`content-type, content-language, content-script-type`
- Manejo de cookies:
`set-cookie`
- Actualización/Redirección de la página:
`refresh`
- Transiciones de página:
`page-enter, page-exit`
- Control parental:
`pics-label`
- Manejo de la cache:
`cache-control, pragma, last modified, expires`

2.4.3.1. Meta name

- Información general:

- **abstract**

Proporciona un resumen muy corto de la etiqueta "description".

No es utilizada por todos los motores de búsqueda al contrario que description.

- **author**

Proporciona el nombre del autor del documento actual.

Es recomendable indicar el nombre completo con nombre y apellidos.

```
<meta name="author" content="Joel Castro León">
```

(Para páginas desarrolladas por un equipo o para indicar un destinatario de correo diferente al autor se puede utilizar la etiqueta web_author)

- **copyright**

Sirve para indicar que hay información bajo derechos de autor.

```
<meta name="copyright" content="Editorial Garceta 2012">
```

Es recomendable sin embargo, utilizar un icono o enlace fácilmente visible para especificar que el documento está bajo derechos de autor.



Figura 2.1: Copyright

Actividad 2.6:

Visita la página <http://es.creativecommons.org/> y crea una licencia para tu sitio Web con objetivos no comerciales, luego inserta el código en tu página.

- **date**

Sirve para indicar la fecha de creación de la página.

```
<meta name="date" content="24 de Febrero de 2010">
```

- **generator**

Sirve para indicar la herramienta utilizada para construir el sitio web.

```
<meta name="generator" content="KompoZer">
```

- Información específica para buscadores:

- **description**

Sirve para explicar el contenido y objetivos del sitio web.

Esta información es utilizada por los motores de búsqueda para indexar la página, así que resulta importante el texto que coloquemos.

```
<meta name="description" content="Lenguajes de marcas y sistemas de gestión de información. Editorial Garceta. 2012">
```

El texto indicado en la descripción suele aparecer como resultado de la búsqueda realizada por un usuario.

- **distribution**

Se utiliza para indicar el nivel de distribución que queremos para nuestro sitio web.

Niveles: Global (Internet) | Local (reservado para la red local) | IU (Uso interno).

```
<meta name="distribution" content="global">
```

No es recomendable utilizar esta etiqueta, la forma adecuada de restringir el acceso a nuestro sitio es configurar adecuadamente el archivo .htaccess de nuestro servidor http, o bien crear un archivo robots.txt, para que los motores de búsqueda indexen según nuestras preferencias.

- **keywords**

Sirve para indicar una lista de palabras clave acerca de nuestro sitio web.

Es una etiqueta muy útil para que nuestro sitio sea localizable, sin embargo debe utilizarse correctamente. La inclusión de palabras no relacionadas con nuestro sitio con el fin de mejorar el posicionamiento web, se considera spam y puede ser penalizado por los motores de búsqueda.

```
<meta name="keywords" content="Ciclos, Formativos,
                                Informática, FP, DAW, DAM, ASIR, SMR">
```

- **locality**

Sirve para indicar al robot de indexación la localidad donde se ubica el sitio web.

```
<meta name="locality" content="Madrid, España">
```

Actividad 2.7:

Incluye estas etiquetas en tu página cambiando el contenido de acuerdo a tu empresa y visualízalas con el navegador Firefox.

- **rating**

Sirve para recomendar el visionado de la página según la edad de un modo similar a las películas en el cine. Existen alternativas mejores como Pics-label.

```
<meta name="rating" content="general">
```

- revisit-after

Sirve para indicar a los motores de búsqueda, que revisiten nuestra página para indexarla de nuevo, ya que ha sido modificada.

```
<meta name="revisit-after" content="3 days">
```

Los motores de búsqueda suelen tener su propio patrón de retorno a las páginas ya indexadas, así que no siempre cumplen el contenido de esta etiqueta.

- robots

Sirve para controlar los robots¹ de los motores de búsqueda.

Los valores posibles son:

| Valor | Utilidad |
|-----------|---|
| index | Indica al robot que indexe esta página |
| follow | Indica al robot que siga los hipervínculos de esta página para indexarlos también. |
| all | Equivale a los dos anteriores juntos. |
| noindex | No indexa la página. |
| nofollow | No sigue los hipervínculos. |
| none | Equivalente a las dos anteriores juntas. |
| noarchive | No permite a los buscadores almacenar la página en caché |
| noodp | No permite usar la descripción oficial de la página que hay en Wikipedia (si la hubiera). |
| noidir | Equivalente a la anterior para Yahoo. |

Tabla 2.2: Valores para el archivo robots

```
<meta name="robots" content="index, nofollow">
```

Además de usar esta etiqueta, los motores de búsqueda se comunican con el sitio Web mediante el archivo robots.txt, que debe crear el administrador del sitio y colocarlo en el directorio raíz.

El archivo robots.txt es de fácil creación, se trata de texto plano que se puede generar con el bloc de notas. Su objetivo primordial es limitar la indexación generalizada de las páginas web de nuestro sitio.

Utiliza unas pocas palabras clave como:

User-agent (robot), **Allow** (permitir), **Disallow** (prohibir)

¹ Robot o araña es un programa de rastreo automático de páginas web, es usado por los buscadores en Internet para actualizar sus grandes bases de datos (por ejemplo: googlebot o msnbot).

Y los valores:

* (todos los robots), / (todos los directorios).

Para evitar directorios o archivos en particular hay que deshabilitarlos por separado.

El archivo se lee de arriba abajo ordenadamente, para cada robot se pueden prohibir los directorios y archivos deseados y cada user-agent es independiente del siguiente.

Ejemplos:

User-agent: *

Allow: /

Permite a todos los buscadores indexar todas las páginas, sería equivalente a no crear el archivo.

User-agent: *

Disallow:

Lo mismo que el anterior.

User-agent: *

Disallow: /imagenes/

Disallow: /js/

Impide a todos los buscadores indexar los directorios: imagenes y js.

User-agent: *

Disallow: /

User-agent: googlebot

Disallow: /

Allow: public

Impide a todos los robots indexar todas las carpetas y permite al robot googlebot indexar solamente el directorio /public.

Hay que tener en cuenta que a pesar de evitar la indexación de una carpeta, si algún recurso es referenciado por un enlace, entonces aparecerá en las respuestas del buscador. Para evitar esto hay que utilizar la metaetiqueta robots con "noindex, nofollow".

Actividad 2.8:

- En tu nuevo sitio crea dos carpetas una llamada "public" y otra "private".
 - Crea un archivo robots.txt que no permita indexar la carpeta private y súbelo al directorio raíz.
 - Inserta una etiqueta meta para que no siga ningún hipervínculo.
-

- Para administración del sitio
 - **google-site-verification**

Sirve para identificar al administrador del sitio en las herramientas de administración de sitios de google. Estas herramientas permiten a un administrador analizar varios parámetros: como su visibilidad, las búsquedas que recibe el sitio, sitemaps, errores, etc.

```
<metaname="google-site-verification"
content="8EYm4qdOa4U0L8tRw9clHsjSbo_JOK0Jjlxaa4OtQAA">
```

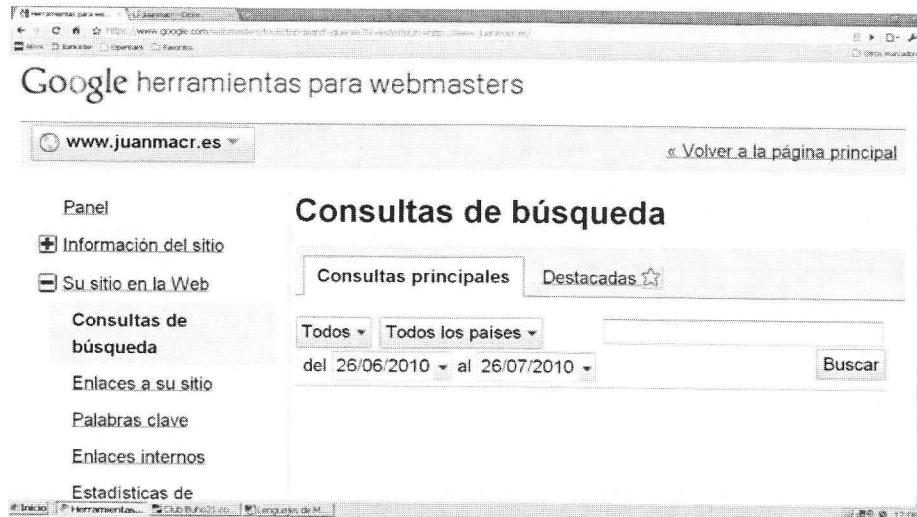


Figura 2.2: Herramientas para webmasters de Google

Actividad 2.9:

- En herramientas para webmasters de Google solicita un código de verificación, mediante la opción "métodos alternativos", insertar etiqueta meta.
- Insértalo en tu página.
- Comprueba que tu sitio ha sido verificado.
- Visita periódicamente el sitio y observa la variación en las estadísticas.

- reply-to

Sirve para indicar una dirección email para contactar con el administrador del sitio.

No es nada recomendable usar esta etiqueta pues puede utilizarse como herramienta de spam, la alternativa es construir un formulario de contacto.

- Dublin-Core

Este proyecto pretende crear un estándar para el uso de metaetiquetas.

```
<meta name="dc.title" content="Enrique Rojo S.A.">
```

2.4.3.2 Meta http-equiv

- Tipo de Contenido:

- Content-type

Sirve para indicar el tipo de los datos contenidos en el cuerpo de la página y en caso de ser texto también indica el juego de caracteres utilizado.

Según el RFC 822 el campo **content-type** se define:

Content-Type := tipo "/" subtipo [";" atributo = valor]

Los tipos MIME (Multipurpose Internet Mail Extensions – Extensiones Multipropósito de Correo de Internet) representan los tipos de formato válidos para un archivo.

| Tipo MIME | Utilidad |
|-------------|---|
| application | Datos binarios o información para programas de correo Subtipos: octect-stream, oda, postscript, x-shockwave-flash, pdf |
| audio | Archivos de sonido, requieren de un periférico para su salida Subtipos: basic, wav, mp3 |
| image | Imágenes, requieren de un periférico para su salida Subtipos: jpeg, gif, x-icon |
| message | Un mensaje encapsulado Subtipos: rfc828, partial, external-body |
| multipart | Datos formados por varias partes con diferentes tipos de datos Subtipos: mixed, alternative, parallel, digest |
| text | Información en forma de texto Subtipos: plain, richtext, html |
| video | Datos de video, requiere software y hardware específico Subtipos: mpeg, avi |

Tabla 2.3: Tipos mime (solo aparecen los más importantes)

```
<meta http-equiv="content-type" content="text/html;
      charset=iso-8859-1">
```

En este caso usamos el juego de caracteres latín nº1 que es adecuado para nuestro idioma. Si queremos hacer más internacional nuestra página, escribiendo caracteres que no pertenecen a nuestro alfabeto, necesitamos el charset utf-8.

El UTF-8 (Unicode Transformation Format – Formato de Codificación Universal) contiene todos los caracteres posibles y cada uno tiene un código diferente. Para nuestros caracteres específicos como los acentos o la ñ debemos utilizar los siguientes códigos:

á = ´ é = é í = í ó = ó ú = ú
ñ = ñ (&=AMPERSAND)

Actividad 2.10:

- Cambia el juego de caracteres a utf-8
 - Inserta en el cuerpo de tu página:
`<p>Cuidado con las esdrújulas</p> <p>Cuidado con las esdrújulas</p>`
 - Observa la diferencia.
-

- Content-language

Sirve para indicar el lenguaje principal del documento como en `lang`.

```
<meta http-equiv="content-language" content="es_es">
```

Puede ser utilizado por los robots para categorizar el sitio web por el idioma.

La lista oficial de códigos de idioma está definida por el IANA.

- content-script-type, content-style-type

Sirve para indicar el lenguaje de programación por defecto usado en el script y el lenguaje de hoja de estilos por defecto en el documento.

Normalmente no se usan porque los navegadores los detectan automáticamente.

• Manejo de cookies:

- set-cookie

Sirve para guardar una cookie en el navegador del usuario.

Una cookie es información personal, como el nombre de usuario o sus preferencias, que sirve para mejorar la navegación cuando revisitamos un sitio web.

Si se especifica la fecha de expiración, la cookie será almacenada hasta el momento indicado, si no se especifica se borrará al cerrar el navegador.

No es lo más adecuado usar esta etiqueta, el manejo de cookies hoy en día se realiza mediante lenguajes de lado servidor como PHP.

- **Actualización/Redirección de la página:**

- **refresh**

Sirve para indicar el tiempo de espera en segundos hasta una actualización o redirección automática de la página.

```
<meta http-equiv="refresh" content="10">
```

Se utiliza habitualmente en páginas que presentan información en tiempo real.

```
<meta http-equiv="refresh"
      content="10;URL=http://otro_sitio.html">
```

Puede utilizarse para redireccionar a otro sitio web.

En el caso de que hayamos cambiado de dominio, es preferible insertar un enlace para que el usuario redirija o el método de redirección 301 (que requiere insertar código en el lado servidor).

- **Transiciones de página:**

- **page-enter, page-exit**

Sirven para realizar transiciones tipo “flash” en la carga de la página y al salir.

```
<meta http-equiv="Page-Enter" content="blendTrans(duration=4)">
<meta http-equiv="Page-Exit" content="revealTrans(duration=1.5,
transition=5)">
```

Solo tienen finalidad estética, no funcionan en todos los navegadores.

También puede hacerse mediante CSS3.

Actividad 2.11:

Inserta las etiquetas anteriores modificando los parámetros de duración y transición, comprueba el resultado.

- **Control parental:**

- **pics-label**

Sirve para indicar la calificación moral del contenido atendiendo a una serie de parámetros, como desnudez, violencia, lenguaje, comportamientos sociales, etc.

Existen varias organizaciones de calificación, una de las más conocidas es el ICRA (Internet Content Rating Association – Asociación para Clasificación de Contenidos de Internet), la cual proporciona un generador gratuito de etiquetas.