

METADATOS Y SITIOS WEB

Anabell Gamboa Comas

Incubador de empresas de base tecnológica del Instituto Politécnico Nacional

anabell@alethia.dgsca.unam.mx

METADATOS Y SITIOS WEB

Resumen:

Los metadatos son datos a cerca de los datos, es decir, los metadatos describen el tipo de información que contiene una página Web. Estos metadatos se guardan en bases de datos que poseen los buscadores y así es como estos identifican rápidamente a las páginas. En la actualidad millones de personas utilizan la Web para buscar y encontrar información de prácticamente cualquier tema. Lo anterior se logra con el uso de diversos buscadores de uso gratuito como yahoo, google, lycos, etc.

Palabras clave: metadatos / web / buscadores / información / tema

METADATA AND WEB SITES

Abstrac:

The metadates are data about the data, that is to say, the metadates describe the type of information that contains a page Web. These metadates keep in data bases that have the finders and thus it is as these identify quickly to the pages. At the present time million people se the Web to look for and to find information of practically any subject. The previous thing is obtained with the use of diverse finders of gratuitous use like yahoo, google, lycos, etc.

Keywords: metadata / web / searchers / information / subject

Inicio

En la actualidad millones de personas utilizan la Web para buscar y encontrar información de prácticamente cualquier tema. Lo anterior se logra con el uso de diversos buscadores de uso gratuito como yahoo, google, lycos, etc. Pero, ¿cómo es que los buscadores localizan las páginas con tan solo teclear unas palabras en su campo de búsqueda? Esto es gracias a la existencia de metadatos; los buscadores pueden generar listas de recursos digitales en la web, los cuales son descritos de forma particular por un conjunto de datos que generalmente están basados en estándares internacionales.

¿Qué son los metadatos?

Los metadatos son datos a cerca de los datos, es decir, los metadatos describen el tipo de información que contiene una página Web.

Estos metadatos se guardan en bases de datos que poseen los buscadores y así es como estos identifican rápidamente a las páginas. Existen diversos métodos de utilizar los metadatos en las páginas Web, algunos ejemplos de usos de metadatos en la informática son los siguientes:

@ Metatags en HTML. Son etiquetas con información sobre el propio documento Web: autor, editor, codificación, etc.

@ Metadatos en la Web semántica. La Web semántica se construye con metadatos RDF en XML.

@ Información en el propio sistema de ficheros. HFS o ReiserFS, por nombrar dos. Se ven completados por los buscadores inteligentes (Beagle o Spotlight) que saben reconocer estos metadatos.

@ Clasificaciones de fotografías. La información EXIF e IPTC almacenada en ellas facilita la catalogación y búsqueda en bancos de imágenes. Programas informáticos como IrfanView, XnView o Picasa por ejemplo, permiten editar y agregar estos datos.

@ Clasificadores de canciones. Contienen metadatos sobre las canciones, bien en MP3 o en CD de audio, en un formato llamado ID3. Por ejemplo iTunes y Rhythmbox.

¿Cómo funcionan los buscadores?

Los motores de búsqueda poseen un programa robot llamado "spider", que recorre las páginas de Internet leyendo los distintos sitios publicados y extrayendo la información necesaria para armar el índice. Por ejemplo los motores como Altavista, utilizan principalmente dos etiquetas:

"Subjects" o "Keywords", para generar los índices de materia y "Description" para elaborar el breve resumen que presenta en la página de resultados.

Altavista dará preferencia a la información suministrada por los Metadatos incrustados en el Header del sitio. De no existir esta información, entonces procederán a extraer la descripción de las primeras líneas de la página y los descriptores del análisis del texto completo.

¿Cómo se utilizan los metadatos en las páginas web?

La aplicación de los metadatos en el diseño de páginas web aporta a la descripción de la forma de las páginas, información sobre su contenido. Incluso se pueden llegar a definir estructuras de datos y las interrelaciones entre los mismos.

Los metadatos están contenidos dentro del código de la página Web y son transparentes para los usuarios, el HTML permite a los autores especificar metadatos de diferentes maneras.

Esto se logra a través de la etiqueta META que puede utilizarse para identificar propiedades de un documento (por ej., el autor, la fecha de caducidad, una lista de palabras clave, etc.) y para asignar valores a esas propiedades. Cabe mencionar que esta etiqueta tendrá que estar dentro de la etiqueta <head> de la página.

Cada etiqueta META especifica una pareja propiedad/valor. El atributo name identifica la propiedad y el atributo content especifica el valor de la propiedad. Por ejemplo, para especificar el autor de un documento, puede utilizarse la etiqueta META como sigue:

```
<META name="Author" content="Anabell Comas">
```

Para especificar las palabras clave y el lenguaje de la página sería de la siguiente manera:

```
<META name="keywords" lang="es" content="vacaciones, Grecia, sol">
```

Metadatos dublín core, qué son y cómo se aplican en las páginas web

El uso de lenguajes para la definición de metadatos estandarizados, tales como XML ó RDF permiten el intercambio de información entre diferentes máquinas, con diferentes sistemas operativos, favoreciendo así la recuperación.

Nacen con este propósito diferentes estándares como Dublin Core Metadata Initiative que pretenden definir una serie de vocabularios de metadatos para describir recursos. De esta forma se puede crear un lenguaje estandarizado que defina recursos de forma internacional. Esto facilita el acceso y la recuperación de información.

El conjunto de Metadatos de la base de Dublín Core es un conjunto de 15 etiquetas propuestas para facilitar la búsqueda de Recursos Electrónicos. Las etiquetas que utilizan metadatos del estándar Dublín Core se caracterizan por usar las iniciales DC. Estas se localizan dentro de la etiqueta <head> </head> en el html de la pagina. El estandar incluye 15 etiquetas básicas con información referente a: título, tema, descripción, tipo de recurso, idioma, documentos relacionados, cobertura, autor, colaborador, derechos de autor, fecha de publicación, tipo, formato y URL¹.

Las etiquetas se encuentran agrupadas en tres bloques:

1. Elementos relacionados principalmente con el contenido del recurso.
2. Elementos relacionados principalmente con el recurso cuando es visto como una propiedad intelectual.
3. Elementos relacionados principalmente con las instancias del recurso.

¹ El total de los elementos del estándar se puede consultar en la siguiente dirección <http://dublincore.org/documents/dces/>
4 -5

Contenido	Propiedad Intelectual	Instancia del recurso
Title	Creator	Date
Subject	Publisher	Type
Description	Contributor	Format
Source	Rights	Identifier
Language		
Relation		
Coverage		

Cada atributo es opcional y puede repetirse. Además, las etiquetas pueden aparecer en cualquier orden. Un ejemplo del uso de metadatos Dublin Core en HTML es el siguiente:

```
<head>
<title>Uso de Metadatos</title>
<meta name="title" content="Los Metadatos">
<meta name="DC.Title" content=" Los Metadatos">
<meta name="keywords" content="metadatos, Web, buscador, Dublin Core">
<meta name="description" content="Esta página contiene información a cerca del uso de metadatos">
<meta name="author" content="Anabell Comas">
<meta name="DC.Creator" content="Anabell Comas">
<meta name="DC.Language" content="es">
<meta name="Language" CONTENT="ES">
<meta http-equiv="DC.Description" content=" Esta página contiene información a cerca del uso de metadatos)">
<meta http-equiv="description" content=" Esta página contiene información a cerca del uso de metadatos">
<meta http-equiv="keywords" content=" metadatos, Web, buscador, Dublin Core ">
<meta http-equiv="title" content=" Los Metadatos">
</head>
```

En la web existen recursos gratuitos que permiten generar el código necesario para crear metadatos. La siguiente dirección proporciona una aplicación que sirve para generar metadatos Dublin Core de tres tipos diferentes (metadatos para HTML ó XHTML, microformatos o para ficheros RDF o XML). Se trata de una aplicación que cuenta con un a interfaz fácil de usar, sólo hay que llenar el formulario con la información correspondiente, presionar el botón "Generar Metadatos" y automáticamente se generará el código correspondiente. El código obtenido se debe insertar en el documento HTML o XML, dentro de sus respectivas etiquetas para que tengan el efecto deseado.

<http://www.webposible.com/utilidades/dublincore-metadata-gen/>

Esta otra aplicación es más sencilla, ya que sólo genera metadatos para HTML y lo realiza a través de un formulario en el que únicamente hay que introducir la información solicitada y dar clic en el botón "Generar Metatags". Para finalmente incluir el código obtenido dentro de las etiquetas <head> y </head> del documento HTML.

<http://www.altas-buscadores.com/metasp.php>