



3. Aplicación de los lenguajes de marcas a la sindicación de contenidos.

Autor	Xerach Casanova
Clase	Lenguajes de Marcas
Fecha	@Mar 29, 2021 9:11 PM

1. Sindicación de contenidos

1.1. Características

Publicación en la web

Sindicación

1.1. Sistemas de gestión de contenidos (CMS)

1.2. Ventajas de la sindicación de contenidos

2. Ámbito de aplicación

3. Tecnologías de creación de canales de contenidos

4. Estructura de los canales de contenidos

4.1. RSS

5. Validación

6. Agregadores

Mapa conceptual

1. Sindicación de contenidos

La redifusión o sindicación de contenidos permite a un sitio utilizar los servicios o contenidos ofertados por otro sitio distinto.

La redifusión web consiste en ofrecer contenido desde una fuente web, cuyo origen esté en otra, proporcionando actualización del mismo.

Los servicios que ofrece la web original junto con los metadatos que tiene asociado, forman los feed o canales de contenidos.

Para leer una fuente o canal hay que suscribirse a ella utilizando un agregador. Esta redifusión se suele realizar bajo licencia de normas de uso o contrato que regule derechos de los contenidos.

Estas fuentes se suelen codificar en XML, pero se puede realizar en cualquier lenguaje que pueda transportar mediante el protocolo HTTP (HyperText Transfer Protocol).

1.1. Características

Publicación en la web

Puede ser visto como un flujo de información que va desde el origen hasta los usuarios que lo leen a través de su navegador, accediendo a una web.

Sindicación

Una web puede ser suministradora de un canal de información (información sindicada), para lograrlo en la cabecera se debe incluir debajo del elemento <title> un enlace al canal de contenidos. Dependiendo de si el canal está hecho con un estándar RSS o con un Atom se utilizan las siguientes líneas:

```
<link rel="alternate" type="application/rss+xml" title="titulo_que_tendrá_el_enlace" href="http://www.misitio.com/fichero.rss" />
<link rel="alternate" type="application/atom+xml" title="titulo_que_tendrá_el_enlace" href="http://www.misitio.com/fichero.atom" />
```

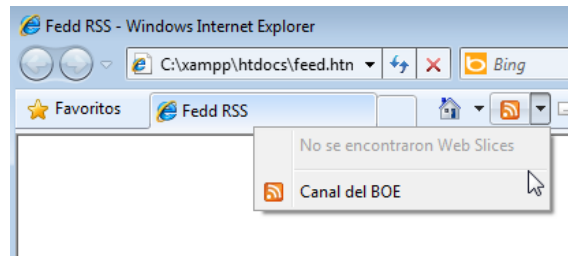
Al vincular un canal, el resultado puede ser poco claro y cambia entre navegadores, por ejemplo:

```

<!DOCTYPE html>
<html>
  <head>
    <title>Fedd RSS</title>
    <link rel="alternate" type="application/rss+xml" title="Canal del BOE" href="https://www.boe.es/rss/boe.php" />
    <meta charset="UTF-8">
  </head>
  <body>
  </body>
</html>

```

El resultado en explorer es:



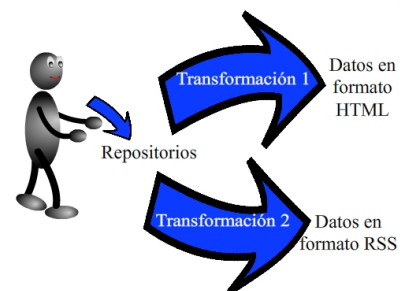
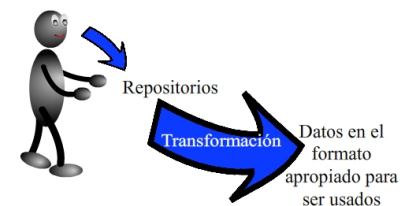
Pero para que el resultado sea más claro, se puede utilizar un vínculo normal, para el que se utiliza normalmente el símbolo de RSS.

1.1. Sistemas de gestión de contenidos (CMS)

Un sistema gestor de contenidos es un programa que permite crear estructuras de soporte para la creación y administración de contenidos en web por parte de los participantes. Disponen de una interfaz que controla una o varias bases de datos donde se aloja el sitio.

El origen de los contenidos es un repositorio y antes de ser servidos al cliente en formato adecuado, se transforman:

- Documento XML → transformación XSLT (lenguaje extensible de transformaciones de hojas de estilo) → documento XHTML.
- Bases de datos → script en Perl → documento HRML
- Texto plano → Página de servidor activo - documento HTML



- Mente del autor → Bloc de notas → documento HTML

Con un CMS de cualquier tipo la transformación puede replicarse. Podemos tener más de una entrada de información y varias salidas (podemos tener tantos ficheros HTML como canales RSS).

1.2. Ventajas de la sindicación de contenidos

- Aumentar el tráfico de la web
- Ayudar a los usuarios a visitarlo frecuentemente
- Favorecer posicionamiento en buscadores
- Ayudar a establecer relaciones entre distintas web dentro de la comunidad

- Permite a otras personas añadir características a los servicios del sitio web (notificaciones de actualizaciones mediante mensajes...)
- Enriquecer internet impulsando la tecnología semántica y fomentar la reutilización.

2. Ámbito de aplicación

La redifusión de contenidos web se puede aplicar a todo tipo de contenidos: texto, audio, videos e imágenes. Se pueden ofrecer contenidos propios en otras webs de forma integrada.

Desde el punto de vista del subscriber, la redifusión permite la actualización profesional. Los usuarios pueden estar al día en temas relacionados con su profesión de sus sitios relevantes.

3. Tecnologías de creación de canales de contenidos

Los estándares más utilizados son:

RSS (Really Simple Syndication), parte de la familia de formatos XML. Se utiliza para:

- Conectar con sistemas de mensajería instantánea.
- Conversión RSS en mensajes de correo electrónico.
- Transformar los enlaces favoritos del navegador en RSS.

Tiene 7 formatos diferentes entre sí:

- RSS 0.90: lo creó Netscape en 1999, se basa en especificación RDF de metadatos, con la intención de que su proyecto My Netscape estuviese formado por titulares de otras webs
- RSS 0.91: versión simplificada lanzada posteriormente. Se detuvo por falta de éxito, pero UserLand Software usó esta versión para desarrollar blogs.
- RSS 1.0: creado a partir del estándar 0.90, más estable y permite definir mayor cantidad de datos que el resto de versiones RSS.
- RSS 2.0: UserLand Software rechazó 1.0 por su complejidad y desarrolló 0.92, 0.93 y 0.94. Todas ellas incompletas y no cumplen todas las normas XML. 2.0 subsana esos errores.

Atom: Fue publicado como un estándar por el grupo de trabajo Atom Publishing Format and Protocol de la IETF en el RFC4287. Se desarrolla como alternativa RSS para evitar confusión por todos los estándares similares, los cuales tenían cierta incompatibilidad entre ellos. Este estándar se desarrolla para convivir con ellos. Es flexible y permite tener mayor control sobre la cantidad de información a representar en los agregadores.

4. Estructura de los canales de contenidos

Para construir un canal de contenido es necesario crear un fichero con extensión rss o atom basado en XML, este se publicará en uno de los directorios de la web en la que se oferta. El fichero consta de:

- Declaración del documento XML y definición de la codificación empleada en el documento, preferiblemente UTF-8.
- Un canal en el que se determina el sitio web asociado a la fuente web a la que se hace referencia en el fichero. Además de su definición está formado por:
 - Secciones: cada una de las cuales es referencia a la web que contiene uno de los servicios que se ofrece. Se pueden incluir tantas como se quieran.

No existen restricciones respecto a la cantidad de canales de contenidos a ofrecer desde una web.

4.1. RSS

El documento RSS incluye como primera línea la declaración del documento XML:

```
<?xml version="1.0" encoding="UTF-8"?>
```

A continuación aparece la etiqueta <rss> que indica que es un documento rss y la versión empleada. Dentro de ella aparece la etiqueta <channel> encargada de describir el feed RSS, con tres elementos hijos obligatorios:

- <title> título del canal
- <link> hiperenlace al canal
- <description> descripción del canal

Y elementos opcionales, tales como:

- <language>
- <category>
- <copyright>

Los canales tienen uno o más artículos (etiqueta <item>). Cada uno cuenta una historia del canal. Tiene 3 hijos obligatorios:

- <title> título del canal
- <link> hiperenlace al canal
- <description> descripción del canal

y elementos opcionales, tales como:

- <author>
- <category>
- <guid> Define un identificador único para el elemento

```
<?xml version="1.0" encoding="UTF-8"?>
<rss>
  <channel>
    <title></title>
    <link></link>
    <description></description>
    <item>
      <title></title>
      <link></link>
      <description></description>
    </item>
  </channel>
</rss>
```

5. Validación

Para validar un documento RSS se le da la dirección del fichero donde se encuentra alojado y comprueba que lo pueden encontrar (URL válida) y que no contiene errores. Una vez validado se obtiene una imagen del tipo XML o RSS de color naranja que se puede incluir en la página principal para enlazar a la dirección del fichero.

Algunos de estos validadores son:

- [FeedValidator](#).
- [W3C Feed Validation Service mediante URL](#).
- [W3C Feed Validation Service mediante código](#).
- [RSS Advisory Board](#).
- [Googletransitdatafeed](#).

6. Agregadores

Agregador o lector de fuentes es una aplicación de software para suscribirse a fuentes en formatos RSS y Atom. Estos agregadores avisan de qué páginas incorporan nuevo contenido desde la última lectura.



En el agregador se debe indicar la dirección a cada archivo fuente RSS o Atom.

Existen varios tipos:

- **Agregadores Web o agregadores en línea.** Aplicaciones que residen en determinados sitios web y que se ejecutan a través de la propia web. Recomendables cuando el usuario no accede desde el mismo ordenador: Feedly, Inoreader o NewsBlur
- **Agregadores de escritorio.** Aplicaciones instaladas en el ordenador. Interfaz gráfica parecida a clientes de correo electrónico, con panel donde se agrupan subscripciones y otro donde se accede a las entradas individuales para su lectura: RSSOwl, QuiteRSS.
- **Complementos de navegador,** disponibles como complementos de navegador web: Awesome RSS o LiveMarks

Mapa conceptual

