DAM

Linguaxes de Marcas e Sistemas de Información

UD4 RSS

PROFESORA | CRISTINA PUGA FDEZ

UNIDADE | UD4

TRIMESTRE | 2

MÓDULO | LMSXI

Tabla de contenido

4.1	Sin	dicación de contenidos	2
4.	1.1	Características	3
4.	1.2	Ventajas de la redifusión de contenidos	5
4.2	Ám	bitos de aplicación	6
4.3	Tec	cnologías de creación de canales de contenidos	7
4.4	Est	ructura de los canales de contenidos	8
4.	4.1	RSS	8
4.5	Val	idación	12
4.6	Agr	regadores	12
ANE	XO I:	Materiales	14

4. RSS

4.1 Sindicación de contenidos

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

Un ejemplo de redifusión podemos encontrarlo en el mercado televisivo. Supongamos una serie de televisión, que es creada por una cadena de televisión. Al principio, sólo la emite esa cadena, en exclusividad. Pero con el paso del tiempo, la vendió a otras cadenas. Estas otras cadenas, al emitir la serie, hacen redifusión.

La **redifusión web** consiste en ofrecer un contenido desde una fuente web, cuyo origen está en otra página web. Se proporciona a los usuarios la actualización del mismo. Los servicios que ofrece el sitio web original, junto con los **metadatos** que tiene asociados en el sitio original, forman los **feed** o **canales de contenidos**.

Para leer una fuente, o canal, hay que suscribirse a ella utilizando un .

La redifusión de contenidos web suele realizarse bajo una licencia de normas de uso, o mediante

un contrato que regule los derechos de los contenidos.

Las suelen codificarse en lenguaje **XML**, aunque es válido hacerlo en cualquier lenguaje que se pueda transportar mediante el (HyperText Transfer Protocol, significa Protocolo de Transferencia de Hipertexto).

Las fuentes generalmente suelen estar codificadas en XML (XML = eXtensible Markup Language, Lenguaje de marcas extensible), aunque es válido en cualquier idioma que pueda transportarse utilizando el protocolo http (HTTP = HyperText Transfer Protocol, Protocolo de transferencia de hipertexto).)

Para leer una fuente, el canal debe estar suscrito a él utilizando un agregador.

Autoevaluación:

La sindicación de contenidos permite...

- a) Que el usuario/a de un sitio web reciba en su correo información sobre cuando se actualiza una web a la que está suscrito/a.
- b) Que el usuario/a de un sitio web pueda acceder a una información o servicio que se encuentra en un sitio web diferente
- c) Los usuarios y usuarias de varios sitios web puedan acceder a varios sitios diferentes desde una misma web.
- d) Que los agregadores lean un canal rss

Respuesta correcta: b)

4.1.1 Características

4.1.1.1 Publicación en la web

Publicar en la web puede ser visto como un flujo de información, que va desde un cierto origen hasta los usuarios y usuarias que la leerán. Podrán hacerlo a través de su navegador, es decir, accediendo a una página web disponible en Internet.

Supongamos que el flujo de información de una publicación tiene su origen en unos ficheros localizados en un ordenador local, codificados en un documento.

Lo que tendremos que hacer, para hacer llegar esa información a los lectores, es subir dichos documentos al directorio adecuado del servidor web que contiene la página.

4.1.1.2 Sindicación

Podemos hacer que una web se convierta en suministradora (origen) de un canal de información, de modo que esta información pueda ser sindicada. Para lograrlo, en la cabecera de la página web hay que incluir, debajo del elemento <title>, un enlace al canal de contenidos.

Para lograrlo, hay que usar una de las dos líneas siguientes, dependiendo de que el canal esté hecho con un estándar o con uno Atom, respectivamente:

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

Al vincular un canal de esta manera, el resultado puede ser poco claro y cambiar entre navegadores. Por ejemplo, para el siguiente documento:

El resultado en el navegador sería:



Es decir, se muestra un menú desplegable con las fuentes vinculadas (si la barra correspondiente está activada).

Para que quede más claro, se puede poner un vínculo normal, para el normalmente se utiliza el <u>símbolo de RSS</u>.

4.1.1.3 Sistemas de Gestión de Contenidos (CMS)

Actualmente es habitual el uso de algún Sistema de Gestión de Contenidos.

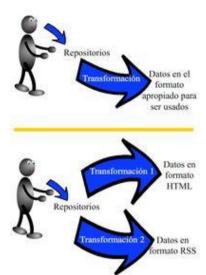
En este caso el origen de los contenidos es un repositorio y, antes de ser servidos al cliente en el formato adecuado, sufren algún tipo de transformación. La parte superior de la figura muestra la estructura del flujo de la información en este caso. Incluso puede haber más de un repositorio.

Esta transformación puede corresponder a uno de los siguientes casos:

- Documento XML -> Transformación XSLT-> Documento XHTML.
- Base de datos -> Script Perl -> documento HTML.
- Texto plano -> Página de servidor activo -> documento HTML.
- Mente del autor -> Bloc de notas -> Documento HTML.

Cuando se usa un CMS de cualquier tipo, la transformación se puede replicar.

Además de tener más de una entrada de información podríamos tener varias salidas. Por ejemplo, podemos generar tanto archivos **HTML** como canales **RSS** como se muestra en parte inferior de la figura.



Autoevaluación:

La única forma de publicar un canal en una web es a través de un fichero HTML

- a) Verdadero
- **b)** Falso

Respuesta correcta: b)

4.1.2 Ventajas de la redifusión de contenidos

Dentro de las ventajas del uso de canales de contenidos de otros propietarios, podemos mencionar:

- Aumentar el tráfico de nuestro sitio web
- Ayudar a que los usuarios y usuarias visiten frecuentemente el sitio web.
- Favorecer el posicionamiento del sitio en buscadores
- Ayudar a establecer relaciones entre distintos sitios web dentro de la comunidad

 Permitir a otras personas añadir características a los servicios del sitio web (por ejemplo, notificaciones de actualizaciones mediante mensajes instantáneos), aunque se necesite de tecnologías adicionales.

Enriquece Internet impulsando la tecnología semántica y fomentando la reutilización.

Autoevaluación:

Selecciona las ventajas de la sindicación de contenidos:

- a) Ayudar al sitio web a aparecer en los primeros lugares de los buscadores
- b) Favorece la distribución de contenidos no aclarando la identidad del sitio
- c) Favorece el trabajo con el significado (semántica) de los datos en lugar de ocuparse sólo de los datos
- d) Permite la generación de páginas web independientes

Respuesta correcta: a) y c)

4.2 Ámbitos de aplicación

La redifusión web no es sólo un fenómeno vinculado a los weblogs, aunque ayudó mucho a su popularización. Siempre se sindicaron contenidos y se compartió todo tipo de información en formato XML.

De esta forma, podemos ofreces contenidos propios para que sean mostrados en otras páginas web de forma integrada, lo que aumenta el valor de la página que muestra el contenido y también nos genera más valor, ya que normalmente la redifusión web siempre enlaza con los contenidos originales.

La redifusión de contenidos web se puede aplicar a todo tipo de contenidos, es decir, texto, audio, videos e imágenes.

Desde el punto de vista de los suscriptores, la redifusión de contenidos permite, entre otras cosas, la actualización profesional. Mediante la suscripción a sitios relevantes, el usuario/a puede estar al día en temas relacionados con su profesión, recibiendo noticias e informaciones en su blog o en su programa agregador de noticias.

Autoevaluación:

Selecciona cuales de las siguientes aplicaciones son ejemplos de redifusión de contenidos:

- a) Blog
- b) Facebook
- c) Google
- d) Youtube

Respuesta correcta: a) y d). Los blogs son redifusores de contenidos de tipo texto y fotos. Youtube es redifusor de videos.

4.3 Tecnologías de creación de canales de contenidos

Los estándares más utilizados se clasifican en dos grupos:

RSS (Really Simple Syndication): es parte de la familia de los formatos XML, desarrollado para

compartir la información que se actualiza con frecuencia entre sitios web. Además, se utiliza en la conexión con sistemas de mensajería instantánea, la conversión de RSS en mensajes de correo electrónico, o la capacidad de transformar los enlaces favoritos del navegador en RSS. Fue desarrollado por tres organizaciones diferentes, lo que dio lugar a siete formatos diferentes entre sí:



- RSS 0.90, es el estándar que creó la empresa Netscape en el año 1999. Se basa en la especificación RDF de metadatos, con la intención de que su proyecto My Netscape estuviese formado por titulares de otras webs.
- RSS 0.91, es la versión simplificada de RSS 0.90 que Netscape lanzó posteriormente.
 El desarrollo de este formato se detuvo por la falta de éxito, aunque la empresa
 UserLand Software decidió utilizar esta versión para desarrollar blogs.
- RSS 1.0, fue creado a partir del estándar RSS 0.90. Es más estable y permite definir una cantidad mayor de datos que el resto de versiones de RSS.
- RSS 2.0, UserLand Software rechazó el estándar RSS 1.0 por considerarlo complejo y continuó el desarrollo del formato RSS 0.91, publicando las versiones 0.92, 0.93 y 0.94. Su sintaxis está incompleta e no cumple todas las normas deXML. El estándar RSS 2.0 y posteriores se publicaron para corregir, entre otros, estos problemas.
- Atom: fue publicado como un estándar propuesto por el grupo de trabajo *Atom Publishing Format and Protocol* (Formato y protocolo de publicación Atom) de la IETF en el RFC4287. Se desarrolló como una alternativa a RSS, con el fin de evitar la confusió creada por la existencia de estándares similares para la sindicación de contenidos, entre los existía cierta incompatibilidad. En lugar de sustituir los estándares existentes, se creó uno nuevo que

convive con ellos. Se caracteriza por la flexibilidad. Atom permite tener un mayor control sobre la cantidad de información que se va a representar en los agregadores.

4.4 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 fichero se publicará en uno de los directorios del sitio web desde el que se oferta.

Estará formado por los siguientes elementos básicos:

- Declaración del documento xml y la definición de la codificación utilizada en el documento.
 Esta última será, preferentemente, UTF-8.
- Un canal en el que se determina el sitio web asociado a la fuente web a la que hace referencia el fichero. Este, además de su propia definición, estará formado por:
 - Secciones: cada una de las cuales es una referencia a la web que contiene uno de los servicios que se van ofrecer. En un canal se pueden incluir tantas secciones como se quiera, lo que hace que un canal de contenido pueda tener un tamaño muy grande si contiene un gran número de enlaces independientes.

No existe ninguna restricción respecto de la cantidad de canales de contenidos que se pueden ofrecer desde un sitio web.

4.4.1 RSS

Para construir un canal de contenidos utilizando el lenguaje RSS, comenzamos creando, con un editor de texto plano, un fichero con extensión .rss

El documento RSS incluye como primera linea la declaración del documento XML, normalmente:

```
<?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 un canal (etiqueta <channel>), que se encarga de describir el feed RSS propiamente dicho. Tiene tres elementos hijos obligatorios:

- <title> Define el título del canal
- link> Define el hiperenlace al canal
- <description> Describe el canal

También hay varios elementos opcionales. Algunos de ellos son:

- <language> Define el idioma del canal
- <category> Define una o más categorías a las que pertenece la fuente
- <copyright>

Cada canal tiene uno o más artículos o secciones (etiqueta <item>), cada uno de los cuales cuenta "una historia" del canal. Tiene tres elementos hijos obligatorios:

- <title> Define el título del artículo
- link> Define el hiperenlace al artículo
- <description> Describe el artículo

También hay varios elementos opcionales. Algunos de ellos son:

- <author> Define el autor del artículo
- <category> Define una o más categorías a las que pertenece la fuente
- <guid> Define un identificador único para el elemento

4.4.1.1 Esquema (con un item)

4.4.1.2 Ejemplo

Como ejemplo, veamos el canal RSS del Boletín Oficial del Estado. En la dirección https://boe.es/rss/boe.php podemos encontrar la información para el último BOE publicado.

```
Triss Wersion="2.0">
- criss version="2.0">
- criss version="2.0"
```

Con el navegador se pueden contraer y expandir los elementos para ver el documento más cómodamente.

```
-<rss version="2.0">
  -<channel>
    <title>BOE - Boletín Oficial del Estado</title>
    k>http://www.boe.es/diario boe/</link>
   -<description>
     Leyes, disposiciones, actos, textos legales y anuncios publicados en la edición de hoy
    </description>
    <language>es-es</language>
    <pubDate>Sat, 21 Dec 2019 00:00:00 +0100</pubDate>
    <lastBuildDate>Sat, 21 Dec 2019 00:00:00 +0100</lastBuildDate>
    <webMaster>webmaster@boe.es</webMaster>
  +<item></item>
  +<item></item>
```

El elemento raíz es **rss**. Tiene un único hijo, **channel**, que define el canal de noticias. Como hijos de este elemento hay varios elementos item, uno por cada noticia publicada en el canal.

Para cada **item**, hay un título (**title**), un vínculo (**link**), una descripción (descripción), una o más categorías (**category**), un elemento guid y una fecha de publicación (pubDate).

El primer elemento **item** es el sumario, como se puede ver en el título, la descripción y la categoría.

En este caso los elementos elemento **guid** y **link** coinciden. Si se accede a esa URL en el navegador, se encuentra, efectivamente, el sumario del BOE para esa fecha.

El resto de elementos **item** se corresponden con los artículos publicados ese día. Tienen dos elementos **category** y los elementos link y guid no son iguales.

Si nos fijamos en el segundo, se trata de la publicación de un acuerdo internacional.

```
-<item>
 -<title>
    Acuerdo entre el Gobierno del Reino de España y la Secretaría de la Convención Marco de las Naciones Unidas sol
    noviembre de 2019.
  </title>
 -<link>
    http://www.boe.es/diario boe/txt.php?id=BOE-A-2019-18359
  </link>
 -<description>
    I. Disposiciones generales - MINISTERIO DE ASUNTOS EXTERIORES, UNIÓN EUROPEA Y COOPERACIÓN - Acuerdo:
  </description>
   <category>I. Disposiciones generales</category>
 -<category>
    MINISTERIO DE ASUNTOS EXTERIORES, UNIÓN EUROPEA Y COOPERACIÓN
   </category>
 -<guid isPermaLink="true">
    http://www.boe.es/boe/dias/2019/12/21/pdfs/BOE-A-2019-18359.pdf
   <pubDate>Sat, 21 Dec 2019 00:00:00 +0100</pubDate>
 </item>
```

Tiene dos elementos **category**, uno para indicar que es una disposición general, y otro para indicar que se trata del Ministerio de Exteriores.

El elemento **link** contiene un vínculo que lleva a una página con el contenido en **HTML**. El elemento **guid** lleva a la versión en PDF del mismo contenido.

Para generar un canal de contenidos en RSS es necesario...

- a) Un fichero XML que además de tener una declaración XML ha de tener otra de RSS.
- b) Un fichero HTML en el que se establecen los enlaces de los diferentes servicios que se sindican.
- c) Dos ficheros XML diferentes, uno para definir el canal y otro para los ítems.
- d) Un fichero RSS en el que hay definido al menos un ítem en el interior del elemento rss.

SOLUCIÓN: a)

4.5 Validación

En internet hay múltiples lugares que dan este servicio.

Para validar un documento RSS con uno de estos validadores, se le da la dirección del fichero donde se encuentra alojado y comprueba que lo pueden encontrar, es decir que la es válida, y que no contiene errores.

Congratulations!



Una vez validado, suelen ofrecer una imagen del tipo "XML" o "RSS", de color naranja por lo general, que se puede incluir en la página principal, para enlazar a la dirección del fichero alojado en su dominio. Así, cuando un visitante pulse sobre este pequeño icono, accederá directamente al contenido actual de la fuente y podrá navegar a través de él a las páginas que más le interesen.

Algunos de estos servicios de validación también ofrecen imágenes que se pueden incluir en la página para que cualquier visitante compruebe que el canal es válido.

Algunos de los validadores que podemos encontrar en Internet son:

- FeedValidator: http://www.feedvalidator.org/
- W3C Feed Validation Service mediante URI: http://validator.w3.org/feed/#validate-by-uri
- W3C Feed Validation Service mediante código:
 https://validator.w3.org/feed/#validate by input
- RSS Advisory Board: https://www.rssboard.org/rss-validator/
- Googletransitdatafeed: https://github.com/google/transitfeed/wiki/FeedValidator

4.6 Agregadores

Caso práctico

María preguntó a Juan, si al trabajar con estas tecnologías de sindicación hay que escribir los ficheros en el bloc de notas o, al igual que al trabajar con HTML, XHTML y XML también existen editores que faciliten la creación de estos ficheros.

Éste respondió que el trabajo puede hacerse con el bloc de notas, pero que, como en el resto de los casos, existen editores que permiten a cualquier persona realizar esa tarea sin dificultad alguna. Además de permitir modificar y crear el documento, estas herramientas tienen más funcionalidades.

Juan continúa explicándoles el último paso para llevar a cabo el proceso de sindicación. Después de tener el fichero fuente validado, es necesario sindicarle.

María pregunta qué pasos implica hacerlo.

Juan le explica que no es más que registrar el fichero en un directorio de canales web.

Félix pregunta en qué consiste y Juan le dice que es un sitio web al que basta darle la URL de la fuente que se quiere sindicar. También les explica que en la mayor parte de los casos se requiere estar registrado en el sitio.

Juan les aclara que aquellos clientes que quieran utilizar este nuevo servicio tendrán que utilizar un agregador de contenidos para poder leer estos canales de información.

Félix se interesa mucho por este tema, ya que habrá que informar a los clientes de cómo instalarlo y utilizarlo.

Juan le tranquiliza diciéndole que es tan sencillo de usar como un gestor de correo y que no es imprescindible instalar la aplicación, ya que se puede utilizar un agregador web, en lugar de uno de escritorio.

¿Qué es un agregador o lector de fuentes?

Es una aplicación de software para suscribirse a fuentes en formatos RSS y Atom. El agregador avisa al usuario o usuaria de qué páginas web han incorporado contenido nuevo desde nuestra última lectura y cuál es ese contenido.

En el agregador hay que indicar la dirección web de cada archivo fuente, ya sea en formato RSS o Atom, para que pueda acceder a sus contenidos, los interprete y los muestre.

Existen varios tipos de agregadores:

 Los agregadores web (o agregadores en línea), son aplicaciones que residen en determinados sitios web y que se ejecutan a través de la propia web. Son recomendables cuando el usuario

o la usuaria no accede siempre a Internet desde el mismo ordenador. Es el caso de Feedly, Inoreader o NewsBlur.

- Los agregadores de escritorio, son aplicaciones que se instalan en el ordenador del usuario o usuaria. Su uso es aconsejable para quienes accedan a Internet siempre desde el mismo ordenador. Su interfaz gráfica suele ser parecida a la de los programas de cliente de correo electrónico, con un panel donde se agrupan las suscripciones, y otro donde se accede a las entradas individuales para su lectura. Algunos ejemplos serían RSSOwl o .QuiteRSS. También podemos descargarnos por ejemplo la versión de escritorio de Feedly.
- Complementos de navegador. También hay agregadores disponibles como complementos de navegador web, como por ejemplo Awesome RSS o Livemarks.

Las siguientes afirmaciones son diferencias entre los directorios RSS y los agregadores de contenidos:

- a) El uso de directorios de contenidos es obligatorio mientras que el uso de los agregadores es opcional.
- b) Los primeros son usados por los creadores de canales de contenidos y los agregadores por los usuarios de los mismos.
- c) Los directorios de contenidos están en un sitio web mientras que los agregadores de contenidos siempre están en un equipo local.
- d) Los directorios de contenidos sólo son utilizados por los creadores de un canal mientras que los agregadores se usan tanto en el proceso de creación como en el de uso.

SOLUCIÓN: b)

ANEXO I: Materiales

I. Textos de apoyo o de referencia

- Wikipedia. http://www.wikipedia.org
- W3school XML RSS: https://www.w3schools.com/xml/xml rss.asp
- W3C Atom: https://validator.w3.org/feed/docs/atom.html

II. Recursos didácticos

- Apuntes en el aula virtual.
- Ordenador personal, con navegador web y conexión a internet.
- Software para elaboración de documentos de texto.