

Aplicación de los lenguajes de marcas a la sindicación de contenidos

Caso práctico

Desde que **María** y **Félix** se asociaron para formar la asesoría legal y empresarial, la empresa ha ido creciendo. Después de lograr compartir la información de sus negocios individuales con la ayuda de **Juan**, su amigo Técnico Superior de Informática, dueño de una empresa de consultoría informática y **Marina**, la trabajadora de esta empresa, y publicar la página web de la empresa, el número de clientes de su negocio aumentó considerablemente, tal y como esperaban.



[Travelin' librarian - www.flickr.com](http://www.flickr.com/photos/travelinlibrarian/) (CC BY-NC-SA)

Con motivo de la cercanía de las fiestas navideñas, **María** y **Félix** invitan a cenar a todos los colaboradores y trabajadores de la empresa, entre ellos se encuentran **Juan** y **Marina**, cuya empresa se ha convertido en la encargada de los asuntos informáticos de la asesoría. Haciendo un repaso de los logros conseguidos con las mejoras informáticas, bromean preguntando a **Juan** cuál es la siguiente tecnología a utilizar para lograr mejorar los servicios que ofrecen a sus clientes. **Juan**, entre risas, les sorprende con su respuesta, comentando la posibilidad de utilizar la web de la empresa para que los clientes puedan informarse de noticias de actualidad que afecten a sus negocios. Esta propuesta les resulta muy atractiva a **María** y **Félix**, que se interesan por el modo de llevarla a cabo.



[Ministerio de Educación y Formación Profesional](http://www.mec.es/). (Dominio público)

Materiales formativos de FP Online propiedad del Ministerio de Educación y Formación Profesional.

[Aviso Legal](#)

1.- Sindicación de contenidos

Caso práctico

Juan les dice que basta con aplicar los lenguajes de marcas a la redifusión de contenidos web para lograr el nuevo servicio. **María** quiere que le explique con un poco más de detalle en qué consiste la redifusión de contenidos web.

Juan comenta que esta tecnología permite utilizar contenidos que ya existen en otras web y ofrecerlos como servicios a través de la propia Web. Siempre, cumpliendo las licencias de las normas de uso de esos contenidos o, si es el caso, respetando las condiciones del contrato que regula los derechos de ese contenido.

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 agregador.

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



Materiales educativos de la CAM (Uso educativo no comercial)

Autoevaluación

La sindicación de contenidos permite...

- Que el usuario o la usuaria de un sitio web reciba en su correo información sobre cuándo se actualiza una web a la que está suscrito o suscrita.
- Que el usuario o la usuaria de un sitio web puedan acceder a una información o servicio que se encuentra en un sitio web diferente.
- A las usuarias y a los usuarios de varios sitios web poder acceder a varios sitios diferentes desde una misma web.
- A los agregadores leer un canal rss.

No es correcta, porque un sitio sindicado no tiene por qué enviar información de sus actualizaciones a los correos de sus usuarios y usuarias.

Muy bien. Sigue por este camino.

No es la respuesta correcta, porque la sindicación permite a un sitio utilizar los servicios ofertados por otro.

La respuesta no es correcta, ya que los agregadores son las herramientas que nos permiten leer un canal de sindicación.

Solución

1. Incorrecto
2. Opción correcta
3. Incorrecto
4. Incorrecto

1.1.- Características

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 HTML.

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.

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.

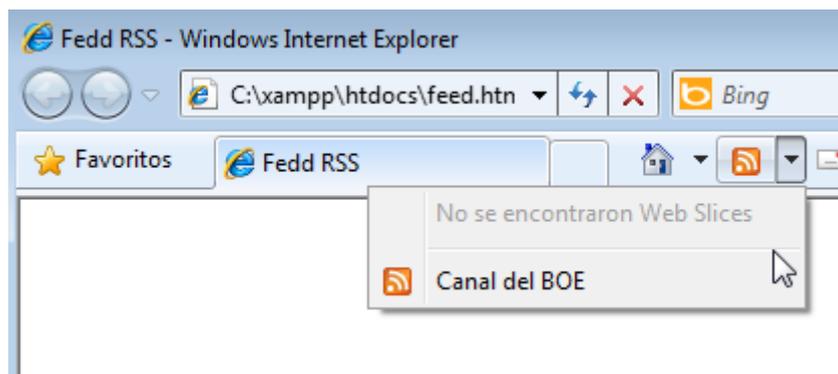
Para lograrlo, hay que usar una de las dos líneas siguientes, dependiendo de que el canal esté hecho con un estándar RSS o con uno **Atom**, respectivamente:

```
1 <link rel="alternate" type="application/rss+xml" title="titulo_que_tendrá_el_enlace" href="https://ejemplo.com/feed" />
2 <link rel="alternate" type="application/atom+xml" title="titulo_que_tendrá_el_enlace" href="https://ejemplo.com/feed" />
```

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

```
1 <!DOCTYPE html>
2 <html>
3   <head>
4     <title>Fedd RSS</title>
5     <link rel="alternate" type="application/rss+xml" title="Canal del BOE" href="https://ejemplo.com/feed" />
6     <meta charset="UTF-8">
7   </head>
8   <body>
9 </body>
10 </html>
```

El resultado en Internet Explorer sería el siguiente:



Materiales educativos de la CAM (Uso educativo NC)

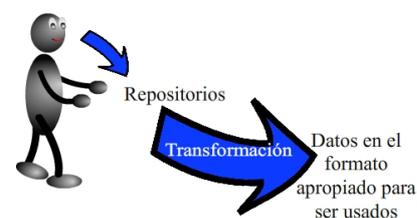
Es decir, se muestra un menú desplegable con las fuentes vinculadas (si la barra correspondiente está activada).

Para que quede más claro, se puede poner un vínculo normal, para el normalmente se utiliza [el símbolo de RSS](#).

Sistemas de Gestión de Contenidos (CMS)

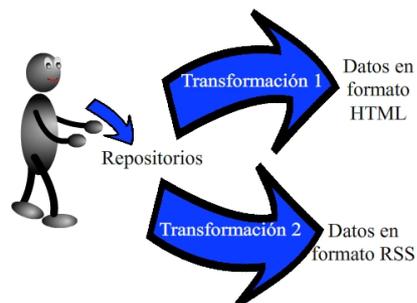
Actualmente es habitual el uso de algún Sistema de Gestión de Contenidos.

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** en **Perl** -> Documento **HTML**.
- ✓ Texto plano -> Página de servidor activo -> Documento **HTML**.
- ✓ Mente del autor -> Bloc de notas -> Documento **HTML**.



[Ana Polo Arozamena con http://opencipart.org/](http://opencipart.org/)
(Uso educativo NC)

Al utilizar un **CMS** de cualquier tipo la transformación puede replicarse. Además de tener más de una entrada de información podríamos tener varias salidas. Por ejemplo, podemos generar tanto ficheros **HTML** como canales **RSS** tal y como se muestra en la parte inferior de la figura.

Autoevaluación

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

- Verdadero.
- Falso.

No es correcta porque puede hacerse la publicación usando algún CMS que parte de un fichero de tipo xml, txt,... y lo transforma luego en un HTML o RSS.

Muy bien. Has captado la idea...

Solución

1. Incorrecto
2. Opción correcta

1.2.- Ventajas de la sindicación de contenidos

¿Cuáles serán las ventajas de utilizar los canales de contenidos de otros propietarios?

- ✓ Aumentar el tráfico de nuestro sitio web.
- ✓ Ayuda a que los usuarios y usuarias visiten frecuentemente el sitio web.
- ✓ Favorece el posicionamiento del sitio en buscadores.
- ✓ Ayuda a establecer relaciones entre distintos sitios web dentro de la comunidad.
- ✓ Permite a otras personas añadir características a los servicios del sitio web (por ejemplo, notificaciones de actualizaciones mediante mensajes instantáneos), aunque se requiera de tecnologías adicionales.
- ✓ Enriquece Internet impulsando la tecnología semántica y fomentando la reutilización.



[geralt - pixabay.com](http://geralt-pixabay.com) (Pixabay License - Todos los logotipos de la Imagen son propiedad de sus respectivas empresas bajo licencia Copyright)

Autoevaluación

Marca las ventajas de la sindicación de contenidos.

- Ayuda al sitio web a aparecer en los primeros lugares de los buscadores.

.....

- Favorece la distribución de contenidos no aclarando la identidad del sitio.

.....

- Favorece el trabajo con el significado de los datos en lugar de ocuparse sólo de los datos.

.....

- Permite la generación de páginas web independientes.

.....

Mostrar retroalimentación

Solución

1. Correcto
2. Incorrecto
3. Correcto
4. Incorrecto

Google.

Youtube.

Mostrar retroalimentación

Solución

1. Correcto
2. Incorrecto
3. Incorrecto
4. Correcto

3.- Tecnologías de creación de canales de contenidos

Caso práctico

Viendo que el formato de la información a distribuir no plantea ningún problema, María se cuestiona si para ofrecer estos servicios de información a sus clientes basta con hacer un enlace a los ficheros que contienen la información utilizando HTML o XHTML.

Juan le informa que para syndicar contenidos hay que utilizar alguno de los estándares de sindicación, los cuales están basados en XML y se agrupan en dos estándares:

- RSS
- Atom

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 para:

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



[re-ality - www.flickr.com](http://re-ality-www.flickr.com) (CC BY)

Ha sido desarrollado por tres organizaciones diferentes, lo que ha dado 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 falta de éxito, aunque la empresa *UserLand Software* decidió usar esta versión para desarrollar blogs.
- ✓ **RSS 1.0**, fue creado a partir del estándar el **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 y no cumplen todas las normas de **XML**. El estándar **RSS 2.0** se publicó para subsanar esos 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 confusión

creada por la existencia de estándares similares para la sindicación de contenidos, entre los que existía cierta incompatibilidad. En lugar de sustituir a los estándares existentes, se creó un nuevo estándar que convive con ellos. Se caracteriza por su flexibilidad. **Atom** permite tener un 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 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 empleada** en el documento. Ésta 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. Éste, 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 a ofrecer. En un canal pueden incluirse tantas secciones como se quiera, lo que hace que un canal de contenido pueda tener un tamaño enorme si contiene un gran número de enlaces independientes.

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

4.1.- RSS

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

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

A continuación aparece la etiqueta **<rss>**. Es declaración **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

Esquema (con un item)

```
1 | <?xml version="1.0" encoding="UTF-8"?>
2 | <rss>
3 |     <channel>
4 |         <title></title>
5 |         <link></link>
6 |         <description></description>
7 |         <item>
8 |             <title></title>
9 |             <link></link>
10 |             <description></description>
```


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 resumen, como se puede ver en el título, la descripción y la categoría.

```
-<item>
  <title>Sumario</title>
  <link>http://www.boe.es/boe/dias/2019/12/21/</link>
  <description>
    Sumario del diario núm. 306, correspondiente al sábado 21 de diciembre de 2019. - Referencia: BOE-S-2019-306 - KBytes: 479
  </description>
  <category>Sumario</category>
  <guid isPermaLink="true">http://www.boe.es/boe/dias/2019/12/21/</guid>
  <pubDate>Sat, 21 Dec 2019 00:00:00 +0100</pubDate>
</item>
```

Materiales educativos de la CAM (Uso educativo NC)

En este caso los elementos elemento **guid** y **link** coinciden. Si se accede a esa URL en el navegador, se encuentra, efectivamente, el resumen 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
  </guid>
  <pubDate>Sat, 21 Dec 2019 00:00:00 +0100</pubDate>
</item>
```

Materiales educativos de la CAM (Uso educativo NC)

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.

Autoevaluación

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

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

Muy bien, es correcto.

No es correcta porque estos enlaces constituyen los ítems del canal de contenidos.

No es la respuesta correcta porque los ítems son parte del elemento channel que define un canal de contenidos.

La respuesta es incorrecta debido a que los ítems han de definirse en el interior del elemento channel.

Solución

1. Opción correcta
2. Incorrecto
3. Incorrecto
4. Incorrecto

5.- Validación

Caso práctico

Juan les explica que una vez generado el fichero con el canal hay que verificar que su codificación es correcta. Para ello no es necesario tener en el equipo local ningún elemento especial, ya que basta con tener una conexión a Internet y entrar en uno de los validadores de fuentes de contenidos.

María se interesa por los datos que hay que darle al validador para que realice el trabajo. **Juan** le contesta que hay dos posibilidades. La más habitual es darle al validador la URL del canal que se quiere validar, pero existen validadores que permiten que se les proporcione el código fuente del fichero.

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 URI es válida, y que no contiene errores.

Congratulations!



This is a valid RSS feed.

[Jorge Castellanos - https://validator.w3.org](https://validator.w3.org) (Nota: El logotipo RSS fue creado por Stephen Horlander - Diseñador de Mozilla para los marcadores de Firefox) (Dominio público - CC0)

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.

Debes conocer

Algunos de los validadores que podemos encontrar en Internet son:

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



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 sindicarlo.

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 syndicar. 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:



[Feedly Inc](#) (Todos los derechos reservados)

- ✓ **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 descargarlos 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*.

Autoevaluación

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

- El uso de directorios de contenidos es obligatorio mientras que el uso de los agregadores es opcional.
- Los primeros son usados por los creadores de canales de contenidos y los agregadores por los usuarios de los mismos.
- Los directorios de contenidos están en un sitio web mientras que los agregadores de contenidos siempre están en un equipo local.
- 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.

Para leer un canal web es obligatorio usar un agregador.

Correcto. Continúa por ese camino.

Incorrecta. Los agregadores web residen en sitios web.

No es correcta. Los agregadores sólo se utilizan al leer los contenidos de un canal.

Solución

1. Incorrecto
2. Opción correcta
3. Incorrecto
4. Incorrecto

7.- Para saber más.

Para saber más

Información adicional:

- [W3Schools XML y RSS](#)
- [W3C Atom](#)