

---

# Sindicación de contenidos

Marcos Ruiz García

2023-11-26 9:00:00 +0100

## Índice

<b>1</b>	<b>Sindicación de contenidos</b>	<b>2</b>
1.1	Características . . . . .	2
1.1.1	Publicación en la web . . . . .	2
1.1.2	Sindicación . . . . .	2
1.1.3	Sistemas de Gestión de Contenidos (CMS) . . . . .	4
1.2	Ventajas de la sindicación de contenidos . . . . .	4
<b>2</b>	<b>Ámbito de aplicación</b>	<b>6</b>
<b>3</b>	<b>Tecnologías de creación de canales de contenidos</b>	<b>6</b>
3.1	RSS . . . . .	6
3.2	Atom . . . . .	7
<b>4</b>	<b>Estructura de los canales de contenidos</b>	<b>7</b>
4.1	RSS . . . . .	8
4.1.1	Ejemplo de RSS . . . . .	9
4.2	Atom . . . . .	11
<b>5</b>	<b>Validación</b>	<b>11</b>
<b>6</b>	<b>¿Qué es un agregador o lector de fuentes?</b>	<b>12</b>
<b>7</b>	<b>Bibliografía</b>	<b>12</b>

# 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 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).

## 1.1 Características

### 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 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.

### 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 RSS o con uno Atom, respectivamente:

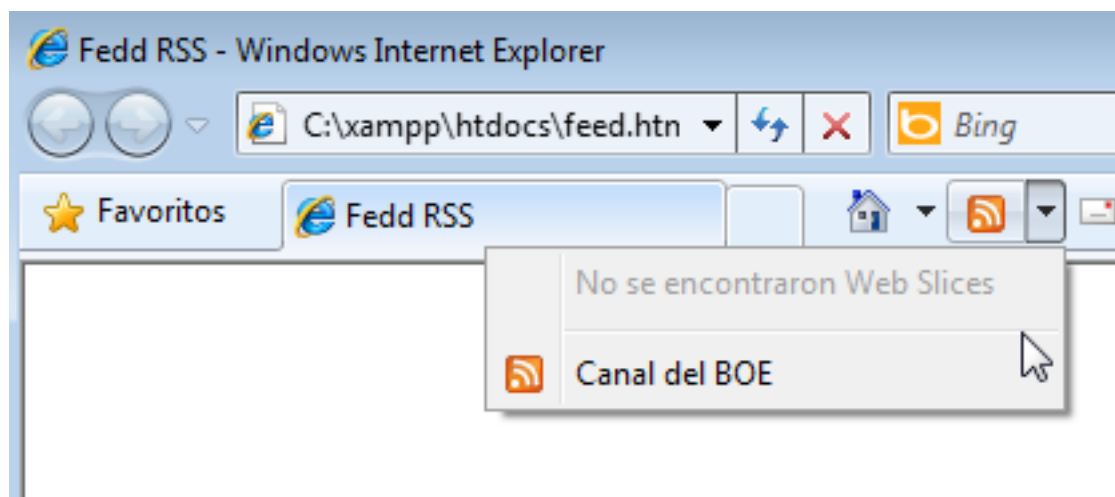
```
1 <link rel="alternate" type="application/rss+xml" title="
  titulo_que_tendrá_el_enlace" href="http://www.misitio.com/fichero.
  rss" />
```

```
1 <link rel="alternate" type="application/atom+xml" title="
  titulo_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:

```
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://www.boe.es/rss/boe.php" />
6     <meta charset="UTF-8">
7   </head>
8   <body>
9   </body>
10  </html>
```

El resultado en Internet Explorer sería el siguiente:



**Figura 1:** Imagen con Canal RSS suscrito en Internet Explorer

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.

### 1.1.3 Sistemas de Gestión de Contenidos (CMS)

Ilustración que muestra el flujo de la información desde la mente del creador, hasta el formato que verá el cliente pasando por el repositorio. Dado que se supone que se utiliza un CMS, la información inicial está localizada en los repositorios.

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.

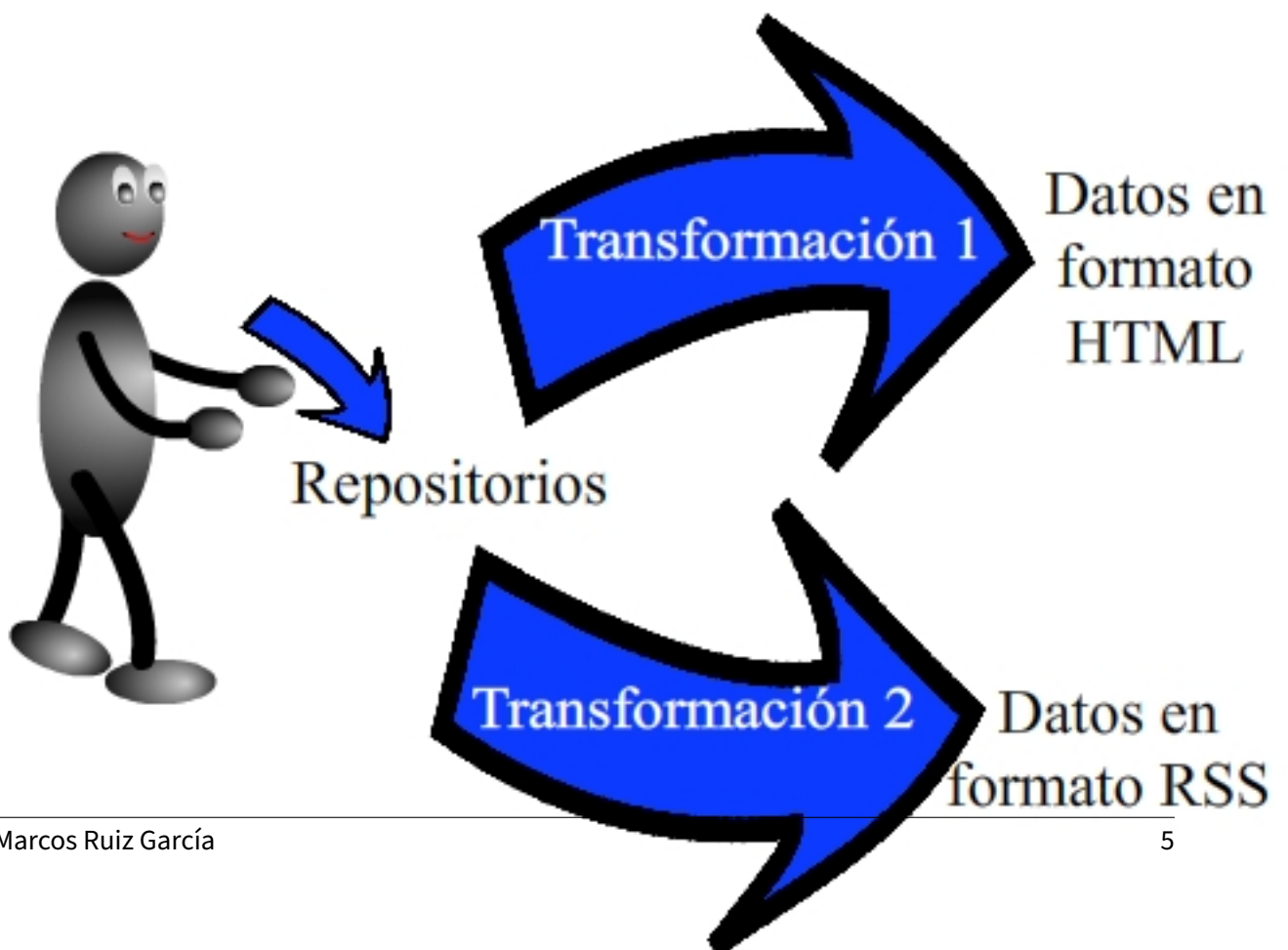
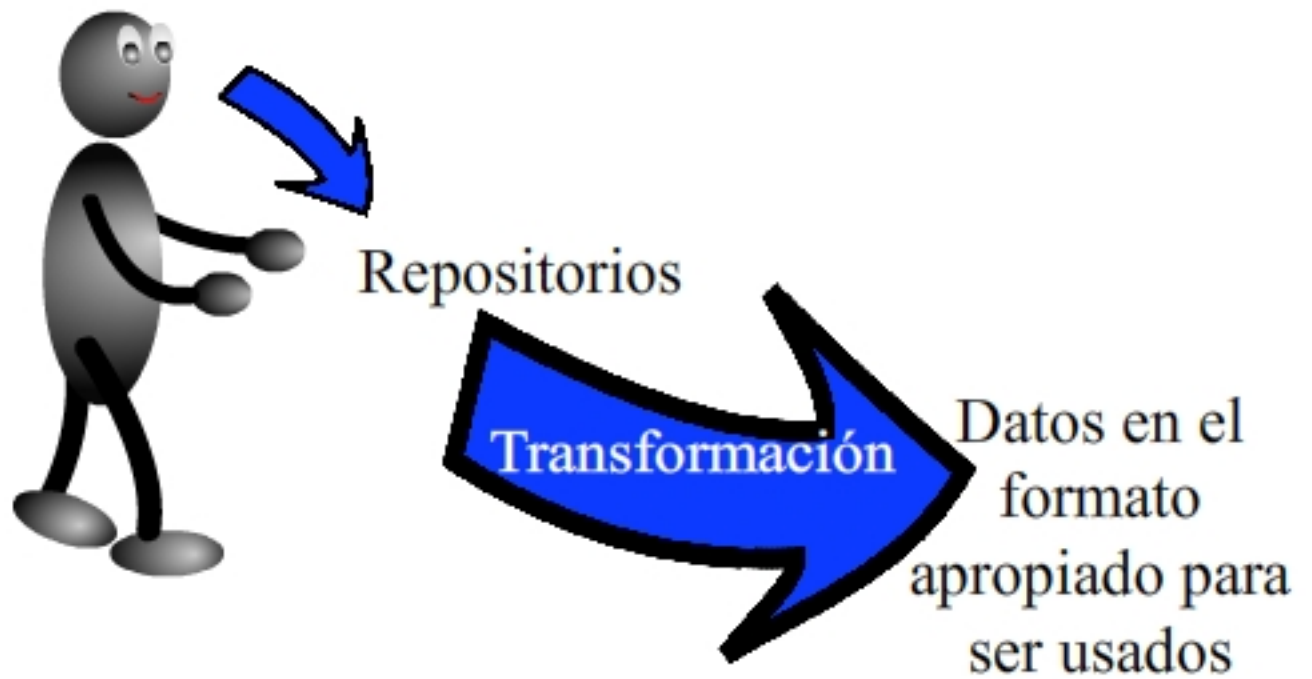
Al utilizar un CMS de cualquier tipo la transformación puede replicarse. Además de tener más de una entrada de información podríamos tener varias salidas. Por ejemplo, podemos generar tanto ficheros HTML como canales RSS tal y como se muestra en la parte inferior de la figura.

## 1.2 Ventajas de la sindicación de contenidos

Las ventajas de utilizar los canales de contenidos de otros propietarios son:

- Aumentar el tráfico de nuestro sitio web.
- Ayuda a que los usuarios 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.

Pregunta: Nombra tres ventajas de utilizar los canales de contenidos de otros propietarios



## 2 Ámbito de aplicación

La redifusión web no es sólo un fenómeno vinculado a los weblogs, aunque ha ayudado mucho a su popularización. Siempre se han sindicado contenidos y se ha compartido todo tipo de información en formato XML.

De esta forma podemos ofrecer 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 puede aplicarse a todo tipo de contenidos, es decir, texto, audio, vídeos 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 puede estar al día en temas relacionados con su profesión, recibiendo las noticias e informaciones en su blog o en su programa agregador de noticias.

## 3 Tecnologías de creación de canales de contenidos

Los estándares más utilizados se clasifican en dos grupos:

- RSS
- Atom

### 3.1 RSS

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.

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.

### 3.2 Atom

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.

Pregunta: ¿Qué significa RFC y para qué se utiliza?

## 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:

```
1 <?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 hipervínculo 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 hipervínculo 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>
11 </item>
12 </channel>
13 </rss>
```

Ejemplo (con un item):

```
1 <?xml version="1.0" encoding="UTF-8" ?>
2 <rss version="2.0">
3
4 <channel>
5   <title>W3Schools Home Page</title>
6   <link>https://www.w3schools.com</link>
7   <description>Free web building tutorials</description>
8   <item>
9     <title>RSS Tutorial</title>
10    <link>https://www.w3schools.com/xml/xml_rss.asp</link>
11    <description>New RSS tutorial on W3Schools</description>
12  </item>
13 </channel>
```

#### 4.1.1 Ejemplo de RSS

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.

```
1 <rss version="2.0">
2   <channel>
3     <title>BOE - Boletín Oficial del Estado</title>
4     <link>http://www.boe.es/diario_boe/</link>
5     <description>Leyes, disposiciones, actos, textos legales y anuncios
6       publicados en la edición de
7       hoy</description>
8     <language>es-es</language>
9     <pubDate>Mon, 27 Nov 2023 00:00:00 +0100</pubDate>
10    <lastBuildDate>Mon, 27 Nov 2023 00:00:00 +0100</lastBuildDate>
11    <webMaster>webmaster@boe.es</webMaster>
12    <item>
13      <title>Sumario</title>
14      <link>http://www.boe.es/boe/dias/2023/11/27/</link>
15      <description>Sumario del diario núm. 283, correspondiente al
16        lunes 27 de noviembre de 2023. -
17        Referencia: BOE-S-2023-283 - KBytes: 453</description>
18      <category>Sumario</category>
```

```

17      <guid isPermaLink="true">http://www.boe.es/boe/dias/2023/11/27/</
18      <pubDate>Mon, 27 Nov 2023 00:00:00 +0100</pubDate>
19      </item>
20      <item>
21          <title>Resolución de 20 de noviembre de 2023, de la Dirección
22          General de Política Energética y
23          Minas, por la que se modifica el Plan de actuación invernal
24          para la operación del sistema
25          gasista.</title>
26          <link>http://www.boe.es/diario_boe/txt.php?id=BOE-A-2023-23990</
27          link>
28          <description>I. Disposiciones generales - MINISTERIO PARA LA
29          TRANSICIÓN ECOLÓGICA Y EL RETO
30          DEMOGRÁFICO - Sistema gasista - Referencia: BOE-A-2023-23990 -
31          KBytes: 204 - Páginas: 3</description>
32          <category>I. Disposiciones generales</category>
33          <category>MINISTERIO PARA LA TRANSICIÓN ECOLÓGICA Y EL RETO
34          DEMOGRÁFICO</category>
35          <guid isPermaLink="true">http://www.boe.es/boe/dias/2023/11/27/
36          pdfs/BOE-A-2023-23990.pdf</guid>
37          <pubDate>Mon, 27 Nov 2023 00:00:00 +0100</pubDate>
38      </item>
39      <!-- Muchísimos más elementos item-->
40  </channel>
41 </rss>

```

Pregunta: ¿Qué es el BOE?

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.

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.

## 4.2 Atom

Un ejemplo de un documento en el formato de distribución Atom:

```
1 <?xml version="1.0" encoding="utf-8"?>
2 <feed xmlns="http://www.w3.org/2005/Atom">
3   <title>Example Feed</title>
4   <subtitle>A subtitle.</subtitle>
5   <link href="http://example.org/feed/" rel="self" />
6   <link href="http://example.org/" />
7   <id>urn:uuid:60a76c80-d399-11d9-b91C-0003939e0af6</id>
8   <updated>2003-12-13T18:30:02Z</updated>
9   <entry>
10    <title>Atom-Powered Robots Run Amok</title>
11    <link href="http://example.org/2003/12/13/atom03" />
12    <link rel="alternate" type="text/html" href="http://example.org/2003/12/13/atom03.html"/>
13    <link rel="edit" href="http://example.org/2003/12/13/atom03/edit"/>
14    <id>urn:uuid:1225c695-cfb8-4ebb-aaaa-80da344efa6a</id>
15    <published>2003-11-09T17:23:02Z</published>
16    <updated>2003-12-13T18:30:02Z</updated>
17    <summary>Some text.</summary>
18    <content type="xhtml">
19      <div xmlns="http://www.w3.org/1999/xhtml">
20        <p>This is the entry content.</p>
21      </div>
22    </content>
23    <author>
24      <name>John Doe</name>
25    </author>
26  </entry>
27 </feed>
```

## 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 URI es válida, y que no contiene errores.

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.

Pregunta: ¿Sabrías decir el nombre de un validador RSS?

## 6 ¿Qué es un agregador o lector de fuentes?

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 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 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. 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.

## 7 Bibliografía

- <https://es.wikipedia.org/wiki/RSS>
- [https://es.wikipedia.org/wiki/Atom\\_\(formato\\_de\\_redifusi%C3%B3n\)](https://es.wikipedia.org/wiki/Atom_(formato_de_redifusi%C3%B3n))
- [https://www.w3schools.com/xml/xml\\_rss.asp](https://www.w3schools.com/xml/xml_rss.asp)
- <https://validator.w3.org/feed/docs/atom.html>