

UD7: Sindicación de contenidos

Lenguajes de
marcas y sistemas
de gestión de
información

Redifusión /
Sindicación de
contenidos



Redifusión/Sindicación de contenidos

- La difusión 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 vende a otras cadenas. Estas otras cadenas, al emitir la serie, hacen redifusión.

Redifusión/Sindicación de contenidos

- La redifusión web consiste en ofrecer un contenido desde una fuente web, cuyo contenido original 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 usando un **agregador (también llamado recopilador de contenidos, agregador de notificaciones o lector RSS)**.

Redifusión/Sindicación de contenidos

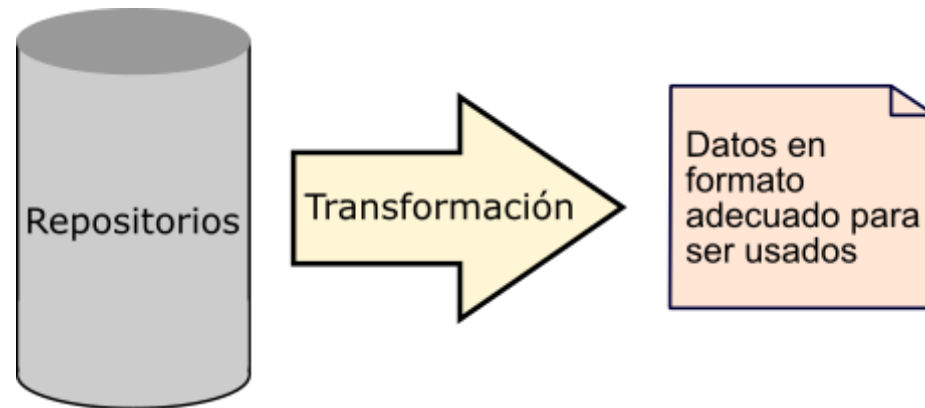
- La redifusión de contenidos web suele realizarse bajo una licencia de normas de uso, o mediante un contrato que regule los derechos de los contenidos.
- Las fuentes suelen codificarse en lenguaje XML, aunque es válido hacerlo en cualquier lenguaje que pueda transportar mediante el protocolo HTTP.

Flujo de la información

- De forma general, la publicación de un contenido web se puede entender como un flujo de información, desde un origen hasta un destino:
 - **Origen:** la página web.
 - **Destino:** los usuarios que leerán la información. Habitualmente a través del navegador.
- 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. Para que los lectores puedan acceder a los documentos, previamente se deben subir a un servidor web que sea accesible desde Internet.

Flujo de la información

- En la actualidad, el uso de Sistemas de Gestión de Contenidos (CMS) está muy extendido para la gestión de ficheros en servidores web y pueden ser una fuente de sindicación de contenidos.
- En este caso, el origen de los contenidos es un repositorio y, antes de ser servidos al cliente en el formato adecuado, se les aplica algún tipo de transformación.



Flujo de la información

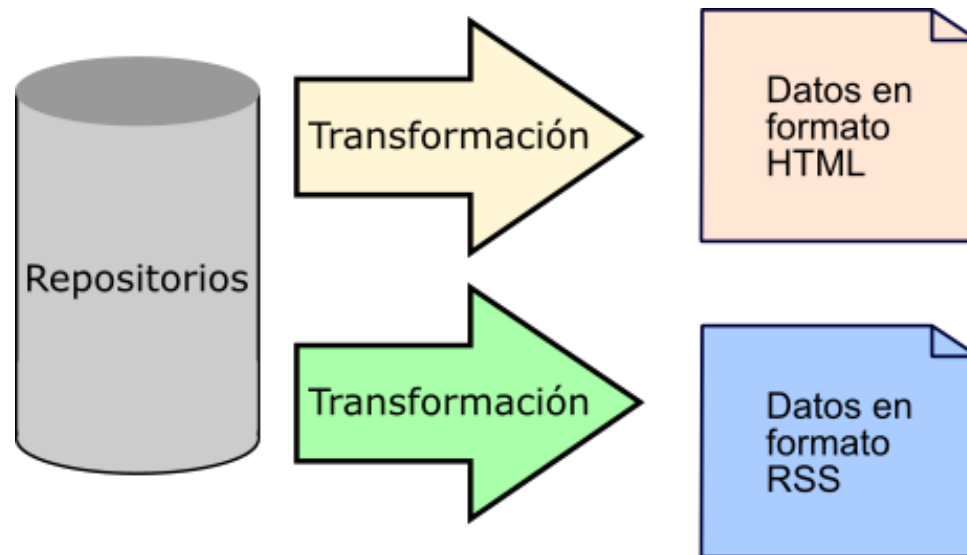
A modo de ejemplo la transformación puede ser:

- Base de datos -> Script -> Documento HTML
- Texto plano -> Página de servidor activo -> Documento HTML
- Documento XML -> Transformación XSLT -> Documento HTML
- Incluso puede incorporarse más de un repositorio a la transformación



Flujo de la información

- Al utilizar un CMS, la transformación puede replicarse.
- De esta forma, además de tener más de una entrada de información, podríamos tener varias salidas. Por ejemplo, podemos generar tanto ficheros HTML como canales RSS.



Estándares



Estándares

- Existen varios estándares para la creación de un canal.
- Todos ellos, en esencia, son **ficheros de texto** (ficheros de texto plano) con la **información estructurada** mediante un **lenguaje de marcas** (como XML o JSON) y que siguen unas reglas definidas en un estándar para la estructuración de la información.
- Veremos los estándares RSS, Atom y JSON Feed.

RSS



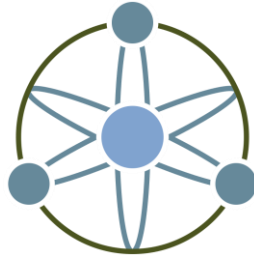
- El estándar RSS (Really Simple Syndication) es el estándar más popular.
- Hace uso del lenguaje XML para estructurar la información.

Ejemplo RSS

```
<?xml version="1.0" encoding="UTF-8" ?>
<rss version="2.0">
<channel>
  <title>RSS Title</title>
  <description>This is an example of an RSS feed</description>
  <link>http://www.example.com/main.html</link>
  <copyright>2020 Example.com All rights reserved</copyright>
  <lastBuildDate>Mon, 6 September 2010 00:01:00 +0000</lastBuildDate>
  <pubDate>Sun, 6 September 2009 16:20:00 +0000</pubDate>
  <ttl>1800</ttl>

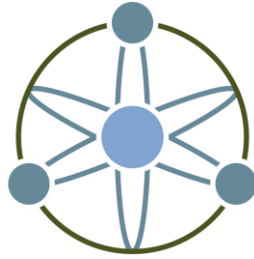
  <item>
    <title>Example entry</title>
    <description>Here is some text containing an interesting description.</description>
    <link>http://www.example.com/blog/post/1</link>
    <guid isPermaLink="false">7bd204c6-1655-4c27-aeee-53f933c5395f</guid>
    <pubDate>Sun, 6 September 2009 16:20:00 +0000</pubDate>
  </item>
</channel>
</rss>
```

Atom



- Atom es un estándar desarrollado como **alternativa a RSS**.
- El objetivo era evitar la confusión que se producía 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.

Atom



- A pesar de todas las mejoras implementadas con respecto a RSS, no tuvo gran adopción debido a que RSS ya estaba implantada.
- Fue publicada como un estándar propuesto por el grupo de trabajo Atom Publishing Format and Protocol de la IETF en el [RFC 4287](#).
- En este documento se indica cómo se debe estructurar la información en el documento usando el formato XML.

Ejemplo Atom

```
<?xml version="1.0" encoding="utf-8"?>
<feed xmlns="http://www.w3.org/2005/Atom">
  <title>Example Feed</title>
  <subtitle>A subtitle.</subtitle>
  <link href="http://example.org/feed/" rel="self" />
  <link href="http://example.org/" />
  <id>urn:uuid:60a76c80-d399-11d9-b91C-0003939e0af6</id>
  <updated>2003-12-13T18:30:02Z</updated>

  <entry>
    <title>Atom-Powered Robots Run Amok</title>
    <link href="http://example.org/2003/12/13/atom03" />
    <link rel="alternate" type="text/html" href="http://example.org/2003/12/13/atom03.html"/>
    <link rel="edit" href="http://example.org/2003/12/13/atom03/edit"/>
    <id>urn:uuid:1225c695-cfb8-4ebb-aaaa-80da344efa6a</id>
    <published>2003-11-09T17:23:02Z</published>
    <updated>2003-12-13T18:30:02Z</updated>
    <summary>Some text.</summary>
    <content type="xhtml">
      <div xmlns="http://www.w3.org/1999/xhtml">
        <p>This is the entry content.</p>
      </div>
    </content>
    <author>
      <name>John Doe</name>
      <email>johndoe@example.com</email>
    </author>
  </entry>
</feed>
```

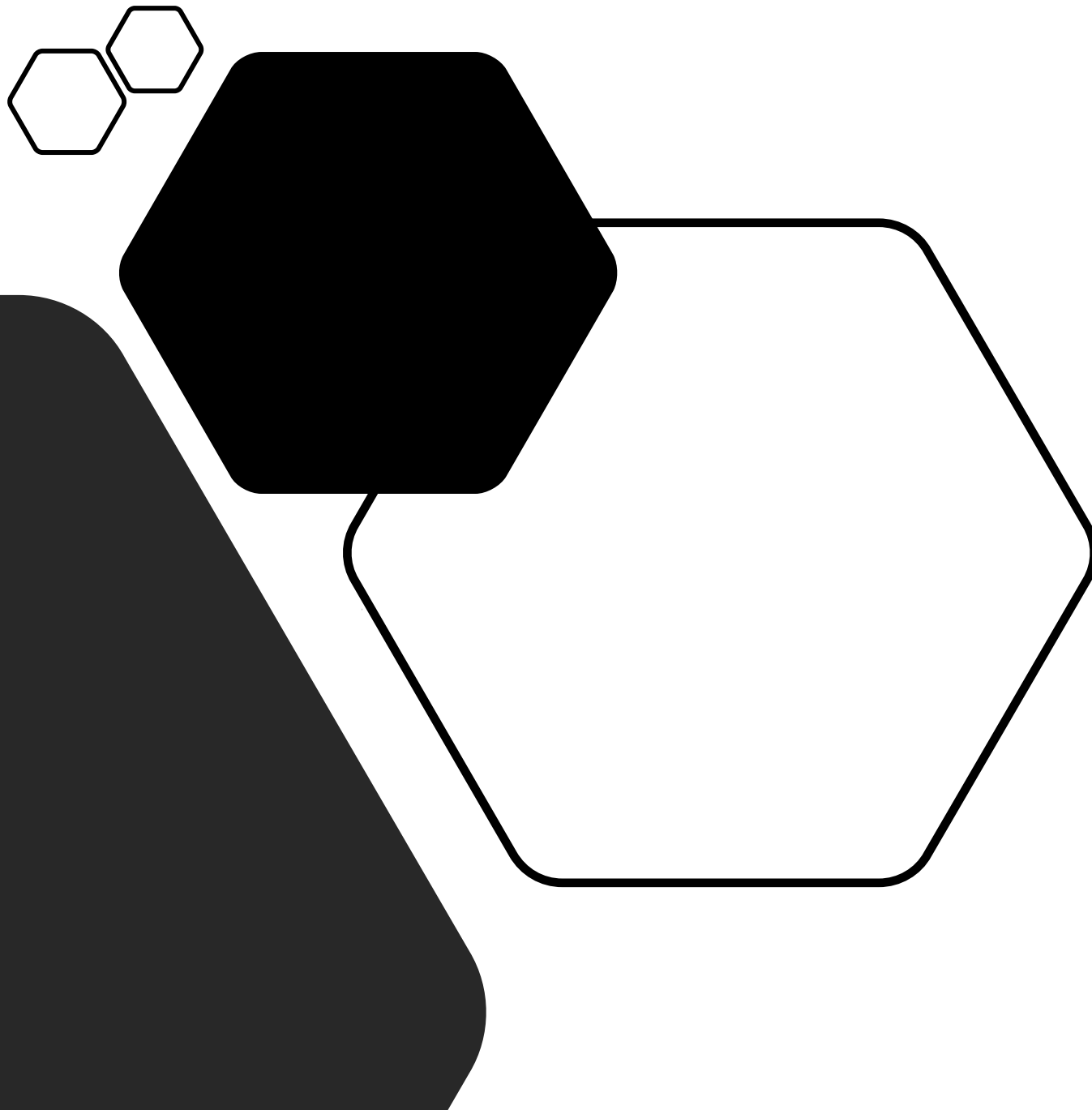

JSON Feed

- JSON Feed es un estándar que usa JSON para la publicación de contenidos. Es el estándar más moderno.

Ejemplo JSON Feed

```
{  
  "version": "https://jsonfeed.org/version/1.1",  
  "title": "My Example Feed",  
  "home_page_url": "https://example.org/",  
  "feed_url": "https://example.org/feed.json",  
  "items": [  
    {  
      "id": "2",  
      "content_text": "This is a second item.",  
      "url": "https://example.org/second-item"  
    },  
    {  
      "id": "1",  
      "content_html": "<p>Hello, world!</p>",  
      "url": "https://example.org/initial-post"  
    }  
  ]  
}
```

Publicación



Publicación

- Podemos hacer que cualquiera 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 (<head>), hay que incluir un enlace al canal de contenidos.
- Para lograrlo, hay que añadir una línea como la siguiente:

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

Publicación

- Si utilizamos otra tecnología, cambiaríamos consecuentemente el valor del atributo type:

```
<!DOCTYPE html>
<html>
  <head>
    <meta charset="UTF-8">
    <title>Feed Atom</title>
    <link rel="alternate" type="application/atom+xml" title="Atom"
href="http://www.misitio.com/fichero.atom" />
  </head>
  <body>
  </body>
</html>
```

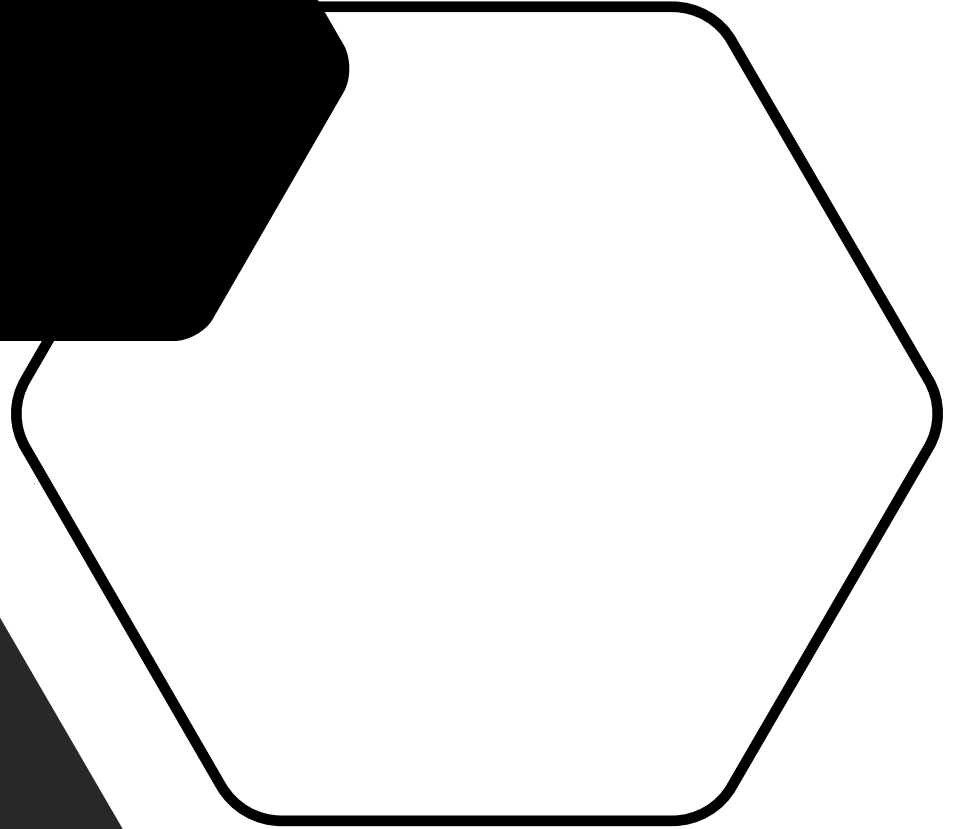
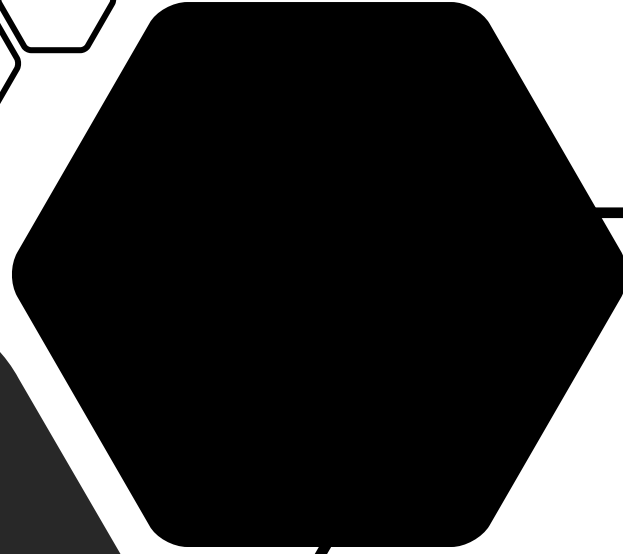
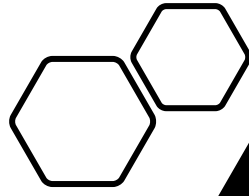
Ámbitos de aplicación

- La redifusión web no es sólo un fenómeno vinculado a los blogs o webs de periódicos, 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.

Ámbitos de aplicación

- La redifusión de contenidos web puede aplicarse a **todo tipo de contenidos**: texto, audio, vídeos, imágenes, etc.
- 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.

RSS



RSS

- RSS 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.

Documento RSS

- Para construir un canal de contenido, es necesario crear un fichero basado en XML.
- Este fichero, que tendrá extensión .rss, se publicará en uno de los directorios del sitio web desde el que se oferta.

Documento RSS

```
<?xml version="1.0" encoding="UTF-8" ?>
<rss version="2.0">
<channel>
  <title>RSS Title</title>
  <description>This is an example of an RSS feed</description>
  <link>http://www.example.com/main.html</link>
  <copyright>2020 Example.com All rights reserved</copyright>
  <lastBuildDate>Mon, 6 September 2010 00:01:00 +0000</lastBuildDate>
  <pubDate>Sun, 6 September 2009 16:20:00 +0000</pubDate>
  <ttl>1800</ttl>

  <item>
    <title>Example entry</title>
    <description>Here is some text containing an interesting description.</description>
    <link>http://www.example.com/blog/post/1</link>
    <guid isPermaLink="false">7bd204c6-1655-4c27-aeee-53f933c5395f</guid>
    <pubDate>Sun, 6 September 2009 16:20:00 +0000</pubDate>
  </item>
</channel>
</rss>
```

Documento RSS: Estructura

- Los elementos básicos que forman parte de un documento RSS son los siguientes:
 - Cabecera del documento.
 - Canal o canales.
 - Secciones.
- La cabecera es la primera línea del documento, que consiste en una declaración xml estándar:

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

Documento RSS: Estructura

- Canal:

- A continuación de la cabecera aparece la etiqueta `<rss>`. La declaración RSS indica que es un documento RSS y la versión usada.

`<rss version="2.0">`

- El elemento `<rss>` es el elemento raíz de un documento RSS.
- Dentro de ella aparece un canal (`<channel>`), que se encarga de describir el feed RSS propiamente dicho. Tiene 3 elementos hijos **obligatorios**.

Elemento	Descripción
<code><title></code>	Título del canal.
<code><description></code>	Descripción del canal.
<code><link></code>	Enlace del canal.

Documento RSS: Estructura

- También hay varios elementos opcionales:

Elemento	Descripción
<language>	Define el idioma del canal.
<category>	Define una o más categorías a las que pertenece la fuente.
<copyright>	Indica quién posee los derechos de autor.

Documento RSS: Estructura

- El **canal** es el que determina el sitio web asociado a la fuente web a la que hace referencia el fichero.
- No existe ninguna restricción respecto a la **cantidad de canales de contenidos** que se pueden ofrecer desde un sitio web.
- Secciones:
 - Un canal está formado por secciones o artículos. Cada una de las secciones 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 desee, lo que hace que un canal de contenido pueda tener un tamaño enorme si contiene un gran número de enlaces independientes.

Documento RSS: Estructura

- Una sección se define con el elemento <item> y tiene 3 elementos hijos obligatorios:

Elemento	Descripción
<title>	Título del artículo.
<description>	Descripción del artículo.
<link>	Enlace del artículo.

- También hay varios elementos opcionales. Algunos de ellos son:

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

Validación

- En Internet podemos encontrar múltiples sitios web que ofrecen este servicio:
 - <https://www.rssboard.org/rss-validator/>
 - <https://www.feedvalidator.org/>
 - https://validator.w3.org/feed/#validate_by_uri
- 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.

Validación

- 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.



Validación

- 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.



Agregadores



Agregadores

- 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:
 - Aplicaciones web.
 - Aplicaciones nativas.
 - Complementos del navegador.

Agregadores

- **Aplicaciones web:**

- Las aplicaciones 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 dispositivo. Algunos ejemplos son [Feedly](#), [Inoreader](#) o [NewsBlur](#).

- **Aplicaciones nativas:**

- Son aplicaciones que se instalan en el **ordenador** o **dispositivo móvil**. Su uso es aconsejable para quienes accedan a Internet siempre desde el mismo equipo. 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 nativa de [Feedly](#) para Android o iOS.

Agregadores

- **Complementos de navegador:**

- Desde que los navegadores no tienen soporte nativo, la alternativa es usar complementos o extensiones, como por ejemplo, [Feedbro](#), [Awesome RSS](#) o [Livemarks](#).

[Threat Intelligence](#)[Market Intelligence](#)[News Reader](#)[Resources](#) [Changelog](#)[LOG IN](#)[GET STARTED FOR FREE](#)

News Reader

For individual readers looking to take control of their news feed

[LEARN MORE](#)

