



Apuntes básicos del Tema 10

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

Contenidos:

1. **Características de la Sindicación de contenidos**
2. **Ventajas de la Sindicación de contenidos**
3. **Ámbitos de aplicación**
4. **Estructura de un sistema de sindicación.**
5. **Funcionalidad y acceso.**
6. **Estándares actuales para sindicación de contenidos: RSS, ATOM.**
7. **Ejemplo de proceso de creación de un fichero RSS.**
8. **Validación.**
9. **Utilización de herramientas: Sistemas de agregación y Directorios de canales de contenidos.**

ANEXO 1: Ejemplos de ficheros RSS y ATOM.

ANEXO 2: Evolución de las distintas versiones



1. Características de la Sindicación de contenidos

Al buscar en un diccionario en inglés, *syndicate* significa "afiliarse a una sindicato", pero además significa "publicar artículos simultáneamente en varios medio de comunicación gestionado por el sindicato al que se pertenece", y también, en EEUU se usa esta palabra para publicar tiras cómicas en diferentes periódicos.

O sea que **sindicación** sería algo así como **suscripción al contenido de un sitio**.

La sindicación de contenidos pretende hacer que parte del contenido esté disponible para su uso a través de otros servicios o canales de información.

El contenido sindicado suele ser simplemente la información más importante del propio contenido, como por ejemplo, titulares, enlaces al contenido, un resumen, etc.

El contenido sindicado se conoce también con el nombre en inglés: **feed**. Que significa "propina" pero también, en éste contexto, feed se traduce como "suministro de datos a un dispositivo".

La Sindicación de Contenidos por medio de canales RSS es un medio para distribuir contenidos en la Web que brinda al usuario la posibilidad de personalizar los contenidos a los que desea acceder.

Se utiliza básicamente para notificar al usuario de los cambios y actualizaciones en un servidor web.

Un ejemplo sencillo: Igual que HTML sirve para escribir páginas en un formato entendible por los navegadores, RSS sirve para enumerar artículos o páginas dentro de un sitio, en un formato que pueden entender programas denominados lectores RSS o agregadores.

En el archivo RSS simplemente están los datos de las novedades del sitio, como el título, fecha de publicación o la descripción. El programa que lee el RSS es encargado de darle estilo o apariencia a los datos que se incluyen en el archivo y presentarlos de una manera atractiva al usuario y de fácil lectura.

Cuando decimos que RSS es un formato basado en XML significa que el archivo RSS está compuesto por una serie de etiquetas definidas que tienen un formato que respeta las reglas de XML.

2. Ventajas de la Sindicación de contenidos

Si por ejemplo, cada día nos conectamos a la red, y visitamos unos cinco blogs favoritos buscando algún artículo interesante, luego visitamos un par de sitios de noticias, y visitamos alguna otra página de nuestro interés, estamos hablando de visitar al menos siete u ocho sitios web, cargarlos en el navegador, buscar el contenido relevante y esperar que descargue todo el texto e imágenes.

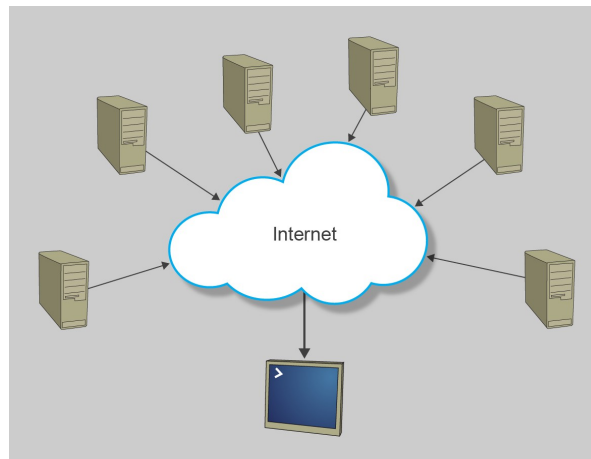
Esta navegación diaria puede llevarnos desde 30 minutos hasta muchas horas. Y podemos encontrarnos con que no hay nada interesante y hemos invertido nuestro tiempo en abrir y cerrar ventanas.

Al utilizar la sindicación de contenidos podemos suscribirnos a los sitios web que queramos y obtener resúmenes de contenido en un solo lugar, darnos de baja cuando queramos. Excepto en servicios de suscripción, no tenemos que dar los datos personales, y es útil, sobre todo, cuando los



contenidos cambian constantemente o de forma regular.

Es la información la que viene a nosotros.



Internet tradicional	La Sindicación de contenidos
Siempre ha facilitado que cualquier usuario participe mediante grupos de noticias y foros de discusión, compartiendo información y conocimiento.	Ha supuesto un avance muy importante para el acceso a la información. Permite la distribución y publicación de un mismo contenido en muchos sitios.
La información está ubicada en portales donde se organiza por temas.	Ofrece una vía mediante la cual podemos seleccionar información valiosa dentro de la gran masa que hoy caracteriza a Internet.
La información está atrapada en repositorios donde hay que acudir para acceder a ella.	La información puede extraerse de sus repositorios originales y situarse en cualquier otro lugar.
El usuario tiene que buscar la información en la web y acceder a sus contenidos de forma manual e individual. Hay que visitar muchas páginas de muchos sitios web para recabar toda la información que nos interesa.	El usuario no tiene que recorrer cada uno de los sitios que le interesan para encontrar la información, sino que recibe, de una forma rápida y eficaz, la información y así ver lo que le puede interesar y descartar lo que no.



3. Ámbitos de aplicación

No se pueden entender las herramientas sociales, sin la sindicación de contenidos.

La potencia que nos ofrece el hecho de podernos syndicar a contenidos etiquetados de estas herramientas colaborativas, para estar al día en todas las fotografías, videos, elementos, que suben a la red los millones de usuarios, sobre un mismo tema, es inmensa.

En España podemos encontrar utilidad de la sindicación en

→ Bibliotecas:

- Biblioteca virtual Miguel de Cervantes
Noticias de la Biblioteca virtual Miguel de Cervantes:
<http://www.cervantesvirtual.com/noticias/noticias.xml>
- Biblioteca de la UOC,
Canales de la Biblioteca Virtual de la UOC:
<http://biblio.uoc.es/rss/esp/ce.xml> (canal de economía)
<http://biblio.uoc.es/rss/esp/gc.xml> (canal de gestión del conocimiento)
<http://biblio.uoc.es/rss/esp/it.xml> (canal de tecnología)

→ Instituciones: Publicaciones, agenda de actividades, departamentos...:

- UCM: <http://www.ucm.es/info/ucmp/cont/news/rssdata/ucm.opml>
- UJI (Canal digital de Tecnología Educativa): <http://cent.uji.es/octeto/rss.xml>
- UGR: <http://www.ugr.es/almacen/canle.xml>
- UCO (Titulares y agenda): <http://www.gestion.uco.es/gestion/comunica/rss/>
- UPSAM <http://www.upsam.com/rss.php>
- UNIOVI (Fd. de C.Económ. y Empr.): <http://www.uniovi.es/econo/Otros/EconoRSS.xml>
- UN (IESE): http://www.iese.edu/aplicaciones/news/xml/rss_es.xml
- UDC (novedades, boletines, notas prensa, ..): <http://www.udc.es/rss/es/fuentes.asp>

→ Medios de comunicación (Algunos requieren suscripción):

- El País <http://www.elpais.es/static/rss/index.html>
- El Mundo: <http://rss.elmundo.es/rss/> (con suscripción)
- La Vanguardia: <http://www.lavanguardia.es/web/public/51142243859.html>
(con suscripción)

→ Organismos/Entidades

- INE: <http://www.ine.es/rss/rssine.rss>
- Barrapunto: <http://backends.barrapunto.com/barrapunto.rss>
- Finanzas.com: <http://noticias.ya.com/rss/>
- EUR-LEX: http://europa.eu.int/eur-lex/es/information/rss_info.html
- Ministerio de Industria, Turismo y Comercio: <http://www2.mityc.es/es-ES/Herramientas/RSS/>

A nivel internacional tenemos algunos ejemplos:

- MIT Libraries (Distribución de sumarios): <http://libraries.mit.edu/help/rss/>
- Princeton University Libraries: <http://diglib.princeton.edu/diglib.rss>
- arXiv.org (áreas temáticas): <http://uk.arxiv.org/help/rss>



- <http://www.library.gsu.edu/news/rss.asp>
- Amazon: <http://www.amazon.com/exec/obidos/subst/xs/syndicate.html/103-3408677-9569446>
- CNN: <http://www.cnn.com/services/rss/>
- BBC: <http://news.bbc.co.uk/1/hi/help/rss/default.stm>

Y si esta herramienta parece potente, se puede multiplicar por mil su capacidad al añadir otros servicios, como por ejemplo: sindicarnos a una búsqueda, es decir, que si nos interesa saber todo lo que aparece sobre Tenis, podemos ir a **Technorati** y formular una búsqueda en todos los post escritos en todos los blogs que filtra dicha compañía (que son millones) sobre ese término.

También existe, para los podcast, la herramienta **PodZinger**, que permite encontrar palabras *dentro* del archivo de audio de un podcast y video podcast. Para realizar las búsquedas y mostrar el contenido en los resultados, el buscador convierte a texto el contenido de voz, y así puede filtrar una gran cantidad de estos elementos multimedia, como si se tratase de un texto.

4. Estructura de un sistema de sindicación.

La estructura de un sistema de sindicación de contenidos tiene relación con el ciclo de vida de los propios contenidos.

Generalmente, el ciclo de vida empieza con el trabajo, creativo e informativo de los autores y termina cuando el contenido llega y es consumido por el usuario.

Para recorrer este camino hay varias opciones de tecnologías y arquitecturas posibles:

1. La más simple es generar el contenido y publicarlo directamente a través de los canales de difusión de páginas web tradicionales, usando HTML, por ejemplo.
2. Otra opción, la más usada actualmente, es servirse de un sistema específico de gestión de contenidos (CMS, Content Management System) que facilita la gestión de los mecanismos de publicación como el desarrollo de contenidos de forma colaborativa. Entre los CMS más habituales de uso libre están Joomla o Drupal, aunque hay otros más.

Cualquiera de estas dos opciones parte de:

- Un contenido, en el caso 1 generado por los autores y en el caso 2 haciendo uso de alguna tecnología (texto plano, bases de datos, documentos XML) y transformar el mismo para presentarlo de forma amigable y estética a los usuarios.
- La salida de esta transformación suele estar compuesta de HTML o XHTML, combinado de CSS, y alguna tecnología adicional (Javascript o Flash).

Pero esta transformación también puede dar lugar simplemente a un conjunto de titulares o metadatos que formateados, usando XML, pueden ser sindicados.

Así, los usuarios tendrán otra interfaz para acceder a los contenidos que les resulten de interés. Esta salida generada en XML directamente no es atractiva para leer los usuarios, por lo que se usan herramientas específicas para transformar dicho contenido sindicado en algo más legible.



5. Funcionalidad y acceso.

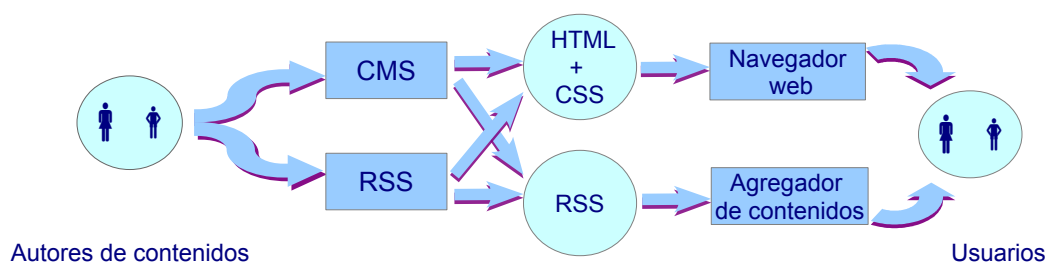
Cuando hablamos de RSS nos referimos a la tecnología para distribución de contenidos de los sitios web. Pero un RSS es un formato de archivo, basado en XML, que sirve para recoger contenidos publicados en páginas web. Los RSS tienen extensión .rss o bien .xml, pero en realidad son un simple archivo de texto donde aparecen referencias a contenidos publicados, en un formato específico, creado a partir de XML.

La función principal es mantener a los usuarios informados sobre lo nuevo que se va publicando en un sitio web sin que éstos tengan que estar conectándose continuamente al portal.

Podemos usar el formato RSS en dos sentidos diferentes:

- * Para recibir información desde otros sitios Web
- * Para ofrecer información desde nuestra propia Web

Según seamos consumidores o creadores de contenidos elegiremos una u otra opción, aunque también podemos utilizar ambas a la vez...



Los canales de sindicación nos permiten syndicar contenidos atendiendo a distintos criterios para adecuarnos más a las necesidades de los usuarios. Por ejemplo, en el caso de un diario de información general, podría crear un canal "internacional", uno "nacional" y uno "deportes", además de uno para todos los titulares.

Además, generalmente, los usuarios se quieren syndicar a más de un canal de generación de contenidos, lo que ha hecho que surjan herramientas que permiten mostrar de forma conjunta los contenidos de distintos canales, permitiendo distintos tipos de organización tanto por temática, como por ejemplo, por fecha de publicación.

Estas herramientas se llaman **agregadores de contenidos** o **lectores de feeds** y proporcionan al usuario un punto de entrada único para acceder a los contenidos que le interesan. Además, permite al usuario no tener que estar accediendo a los distintos canales de forma independiente sino que los propios canales "buscan" al usuario.

Cuanto mayor es la cantidad de canales de contenidos que el usuario quiere acceder de forma habitual, mayor es el tiempo que ahorra usando los agregadores de contenidos.



6. Estándares actuales para sindicación de contenidos: RSS Y ATOM.

1- Identificación estándares.

Cuando un sitio o un blog ofrece un flujo RSS, aparece uno de los siguientes logotipos: Los iconos RSS son testimonio de un canal RSS en el sitio.



Como ampliación, en el anexo 2 se presenta una tabla cronológica que facilita la comprensión de la evolución en el tiempo de los distintos formatos .

2- Algunas particularidades del estándar **RSS**

Hoy en día, el formato más utilizado para sindicación de contenidos es el llamado RSS (Really Simple Syndication).

RSS usa XML para representar el contenido que se va a syndicar y en los últimos años ha evolucionado desde poder agregar unos pocos canales de contenidos a trabajar con una cantidad de información casi ilimitada.

Para escoger la más apropiada para nuestro RSS es necesario conocer las distintas especificaciones, que podemos encontrar en los siguientes sitios:

- RSS 0.91. <http://backend.userland.com/rss091>
- RSS 0.92. <http://backend.userland.com/rss092>
- RSS 1.0. <http://web.resource.org/rss/1.0/spec>
- RSS 2.0. <http://cyber.law.harvard.edu/rss/rss.html>

2- Algunas particularidades del estándar **ATOM**

IBM, Google y otras empresas de hosting alentaron la creación de otra tecnología para vincular contenido, y lo han llamado **Atom**.

No se basa en ninguna versión de RSS, aunque su formato es similar y sirve también para la distribución de contenidos y noticias de sitios web.

Las mejoras que supone respecto a RSS (Un documento Atom puede contener más información y es más consistente que un documento RSS) hacen que su uso se extienda rápidamente a pesar de ser algo más complicado (Ha sido fuertemente alentada por Blogger empresa que pertenece a Google y otra gran cantidad de sitios más).



Los documentos Atom requieren unos cuantos conocimientos más para crearlos partiendo desde cero, ya que la especificación de Atom es complicada (tiene más de 50 páginas).

Podemos encontrar sus especificaciones en :

<http://xml.coverpages.org/draft-ietf-atompub-format-11.txt>

Algunas **diferencias** entre **RSS 2.0** y **Atom** son:

- Los componentes de Atom se llaman **entradas**.
- No hay una etiqueta <channel>.
- Se utilizan etiquetas <**id**> en lugar de etiquetas <guid>.
- Todos los elementos pertenecen al espacio de nombres por defecto de Atom.
- Los vínculos deben incluir los atributos ***rel***, ***type*** y ***href***.



7. Ejemplo de proceso de creación de un fichero RSS.

Los pasos a seguir son:

1. **Declaración** del Tipo de Documento

Se crea un documento XML, que incluye una declaración que lo determina como tal. Por eso, la primera línea de nuestro código RSS será la que define el lenguaje de marcado y el tipo de caracteres que vamos a utilizar.

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

2. **Especificación** RSS

Aquí tendremos que escoger una de las tres especificaciones de RSS que existen. **La versión 2.0 es la que vamos a utilizar en este ejemplo.**

```
<?xml version="1.0" encoding="UTF-8" ?>  
<rss version="2.0">  
</rss>
```

Todas las demás etiquetas que vamos a utilizar para crear nuestro feed deben situarse entre estas dos de <rss>, porque son las que indican que estamos creando un canal RSS.

3. Crear un **canal**, en el que introduciremos los contenidos que queremos mostrar a los usuarios.

Este canal se limita con dos etiquetas <channel> -una de principio y otra de final- a continuación de lo que ya hemos hecho:

```
<?xml version="1.0" encoding="UTF-8" ?>  
<rss version="2.0">  
<channel>  
</channel>  
</rss>
```

Todo feed o canal se compone de dos partes: elementos estáticos y elementos dinámicos.

4. **Elementos estáticos** o fijos.

Existen varios elementos estáticos, pero sólo tres son **obligatorios**:

- Título: <**title**> hace referencia al nombre de nuestro feed o canal, y que debe coincidir con el nombre del sitio web.
- Enlace: <**link**> es la URL del sitio Web
- Descripción: <**description**> Oración que describe el tipo de contenidos que incluye el canal RSS (artículo, receta de cocina, mensaje, nueva sección, lo que sea).



Los elementos estáticos **opcionales** son:

- **language:** Idioma del canal
- **copyright:** Copyright del canal
- **managingEditor:** Nombre y e-mail del responsable del contenido
- **webMaster:** Nombre y e-mail del responsable técnico del servidor
- **pubDate:** Fecha de la última actualización o publicación de ese canal
Mon, 12 Sep 2011 18:37:00 GMT. La fecha es la de la última actualización del feed y debe ser en este formato y la hora siempre debe ser expresada en GMT. Aunque no es un elemento obligatorio es altamente recomendable pues muchos motores de búsqueda lo pueden usar en sus resultados.
- **category:** Categoría a la que pertenece el canal
- **generator:** Software o aplicación utilizado para generar el canal
- **docs:** URL que apunta a la documentación del formato RSS
- **cloud:** Registro para ser notificado de las actualizaciones del canal
- **ttl:** (Time To Live): minutos en que un canal puede permanecer en cache
- **image:** imagen que se desea con el canal
- **textInput:** Cuadro de diálogo para mostrarse con el canal
- **skipHours:** Indicación para los agregadores de las horas a evitar
- **skipDays:** Indicación para los agregadores de los días a evitar

Estas tres líneas de código se escriben entre las dos etiquetas <channel> que hemos ya creado:

```
<?xml version="1.0" encoding="UTF-8" ?>
<rss version="2.0">
  <channel>
    <title>El nombre de nuestro feed RSS</title>
    <link>Dirección web en la que se encuentre nuestro RSS</link>
    <description>Contenido que se ofrece a los usuarios que lean el
    RSS</description>
    <language>ES</language>
    <pubDate>Mon, 19 Mar 2012 13:00:00 GMT</pubDate>
  </channel>
</rss>
```

5. Elementos dinámicos o variables

Los elementos variables de un canal RSS se denominan "ítem" y pueden incluirse varios en un mismo canal (Máximo de 15 ítems por canal).

Cada ítem consta de cinco definiciones de las que tan solo las tres primeras son obligatorias y el resto opcionales (opcionales hay aún algunas más):

Debemos colocar esta información entre las etiquetas <item> </item>.



Se sitúan entre las de <channel> </channel> justo después de los elementos fijos.

Obligatoriamente cada ítem debe contener en nuestro canal **tres elementos variables**, aunque, como en el caso anterior, existen más.

Estos elementos obligatorios vuelven a ser:

- Título: <**title**> Entre cuyas etiquetas escribiremos el título del artículo que queremos divulgar.
- Enlace: <**link**> Que contiene el enlace o URL a la versión completa de la página web
- Descripción: <**description**> donde colocar una corta descripción del elemento (artículo, receta de cocina, mensaje, nueva sección, lo que sea), o sea la dirección Web a la que podemos ir para ampliar esta información.

Y ahora estos elementos describen cada uno de los artículos o informaciones que vamos a ofrecer y cuyo contenido iremos actualizando cada cierto tiempo.

Como definiciones **opcionales**, no obligatorias podríamos añadir estas otras:

- **author**: e-mail del autor del ítem.
- **category**: Donde escribir la categoría o categorías a la que pertenece el elemento.
- **comments**: URL de los comentarios relativos al ítem.
- **enclosure**: Descripción es caso de incluir objetos que forman parte del ítem (audio,...)
- **guid**: identificador unívoco del ítem.
- **pubDate**: Donde indicamos la fecha de la última actualización de ese elemento
- **source**: El canal RSS o fuente del ítem

Hay, por tanto, dos partes bien diferenciadas en el archivo xml: la parte en la que se definen las características del canal (**channel**) por un lado y la zona en la que se define cada artículo (**ítem**).

El código para incluir un ítem en nuestro canal RSS sería el siguiente:

```
<?xml version="1.0" encoding="UTF-8" ?>
<rss version="2.0">
  <channel>
    <title>El nombre de nuestro feed RSS</title>
    <link>Dirección web en la que se encuentre nuestro RSS</link>
    <description>Contenido que vamos a ofrecer a los usuarios</description>
    <item>
      <title>Título del artículo del RSS</title>
      <link>Dirección Web a la que podemos ir para ampliar esta información</link>
      <description>Contenido de esta información</description>
    </item>
  </channel>
</rss>
```

Se deben crear tantos ítems como páginas queramos poner en el archivo, cada una con sus datos propios.



En todos los casos, en los datos ingresados no se puede emplear caracteres especiales como letras ñ, acentos y otros caracteres propios de nuestro idioma.

En la zona en la que se declaran las propiedades del canal se puede añadir una imagen, (por ejemplo el logo de la página web)

A continuación vamos a ver cómo puede ser una **PLANTILLA** para crear un **archivo RSS**.

Guardamos el archivo como : *feed.xml* o cualquier otro nombre, respetando la extensión.

Un feed RSS debe ser validado como documento XML estricto, así que la primera línea del documento es la declaración XML.

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

La parte fija o no variable del fichero xml correspondiente a un rss

```
<?xml version="1.0" encoding="UTF-8" ?>
<rss version="2.0">
  <channel>
    <title>Titulo del canal de RSS</title>
    <link>http://www.mipagina.com/RSS </link>
    <description>Es una pagina web de definición del RSS</description>
```

Y como parte opcional:

```
  <language>es-ES</language>
  <image>
    <title>Título de la Imagen</title>
    <url>http://www.mipagina.com/logotipo-del-rss.png</url>
    <link>http://www.mipagina.com</link>
    <width>90</width>
    <height>36</height>
  </image>
```

Y su respectivas etiquetas de cierre:

```
  </channel>
</rss>
```

La parte variable, la que corresponde a cada noticia o ítem, y un ejemplo puede ser:



```

<item>
  <title>Titulo del primer artículo</title>
  <link>http://www.mipagina.com/ruta/articulo/primero.html</link>
  <description>Breve descripción del contenido del primer
    artículo</description>
</item>
<item>
  <title>Título de un segundo artículo</title>
  <link>http://www.mipagina.com/ruta/articulo/segundo.html</link>
  <description> Breve descripción del contenido del segundo
    artículo</description>
</item>
</channel>
</rss>

```

6. Guardar el RSS y hacerlo visible

Una vez creados el canal y los artículos de nuestro feed guardamos el archivo con extensión .rss o .xml, ya que aunque sea un **documento RSS** está escrito en lenguaje XML.

(La extensión que más se suele utilizar es .rss, pero puede ocurrir que al colgarlo en Internet, algún host no reconozca el archivo. Ante cualquiera de estas situaciones, optamos por guardarlo como **documento XML**).

7. Enlazar el RSS para que sea accesible

Hay que crear un enlace al documento RSS en el sitio web. Por ejemplo el código que podemos incluir en la página puede ser este:

```

<a type="application/rss+xml" href="nombre del canal RSS.xml">  </a>

```

En **type** indicamos el tipo de documento que es.

En la propiedad **href** indicamos el nombre del documento RSS (junto con la ruta de acceso a él, si no se encuentra en el mismo directorio que la página donde estamos poniendo el enlace).

Luego, quien quiera acceder a este contenido deberá indicar la dirección URL donde está el documento a su agregador de RSS.

Para que aparezca el icono de canal rss en la barra de navegación, bastará con buscar entre la etiqueta <head> del código y cambiar la ruta donde se encuentra nuestro archivo correspondiente al canal en la siguiente línea:

```

<link rel="alternate" type="application/rss+xml" title="Mi pagina web" href="xml/rss.xml" />

```

Posteriormente enlazar la opción que se encuentra en la barra de herramientas superior, de la siguiente forma:

```

<a href="xml/rss.xml" onclick="window.open(this.href);return false;">RSS </a>

```



8. Validación.

Como lo que tenemos es un documento escrito en XML, si tiene cualquier error no va a funcionar.

Las especificaciones del W3C indican que las aplicaciones al encontrar cualquier error de sintaxis en el código de un archivo XML, se detengan y no continúen ejecutándose.

Y a diferencia de los archivos HTML (aun con errores siguen funcionando), los XML tienen esa característica, por lo que se hace imprescindible su validación regularmente, más aun si son creados de forma dinámica o modificados de forma automática por varios programas como ocurre con los archivos de fuentes de noticias.

Para validar tenemos dos opciones:

- ➔ Comprobar que el documento está bien formado:
<http://validator.w3.org/>
- ➔ Comprobar que es válido:
<http://validator.w3.org/feed/check.cgi>

Basta con copiar el código de nuestro archivo XML, pegarlo en el cuadro de texto y pulsar el botón "Validar" o también podemos introducir la dirección URL y pulsar el botón "Validar".

Puede ocurrir:

- Que exista cualquier **error**. En ese caso nos muestra un cuadro de alerta informando de la línea donde se encuentra y en que consiste dicho error.
- Que sea **válido**. En ese caso nos aparecerá un mensaje como el siguiente:

Congratulations!



This is a valid RSS feed.

En todos los casos sólo se valida el código XML, no la exactitud del contenido.



9. Utilización de herramientas: Sistemas de agregación y Directorios de canales de contenidos.

En la actualidad, existen en Internet un número elevado de feeds RSS: feeds de noticias, blogs, etc.

Si el usuario quiere poder leer todos estos feeds, necesita una herramienta que le facilite la tarea. Aquí entran en juego los agregadores de contenidos. Son **sistemas de agregación** que permiten mostrar de forma conjunta los contenidos publicados en distintos feeds.

Un **agregador** (de noticias o feeds) es un tipo de aplicación que permite suscribirse a fuentes web en formatos RSS, Atom y otros, para recibir información actualizada de un sitio web y también es llamado agregador de news, lector de noticias, lector de feeds, feed reader, lector RSS.

Funcionamiento básico de un programa agregador:

- Un agregador permite suscribir sitios web que ofrecen un servicio llamado sindicación web. El visitante se suscribe a un sitio web a través de un enlace. El programa agregador se encarga de verificar dicho enlace y descarga (o solo notifica) información de ese sitio web.
- Esta información suele ser un resumen (en títulos, descripciones y direcciones web) del contenido del sitio web y sus actualizaciones.
- El agregador concentra toda la información en un único lugar y el usuario recibe las actualizaciones de sus sitios web preferidos de forma rápida y homogénea, sin tener que visitar cada uno de esos sitios web.
- El programa agregador puede ser programado para que periódicamente revise si hay nuevo contenido a descargar.
- Además los agregadores pueden filtrar, ordenar y buscar entre los múltiples feeds que llegan.

Tipos de Programas o servicios agregadores

1. Agregadores de escritorio. Son programas que se instalan. Hay que descargarse el programa e instalarlo en el ordenador.

Algunos son:

- Para Windows:

- RssReader.
- FeedDemon.
- FeedReader.

- Para GNU/Linux:

- Liferea.
- Evolution.
- Syndicator.
- Straw.

- Para Mac:

- NetNewsWire .

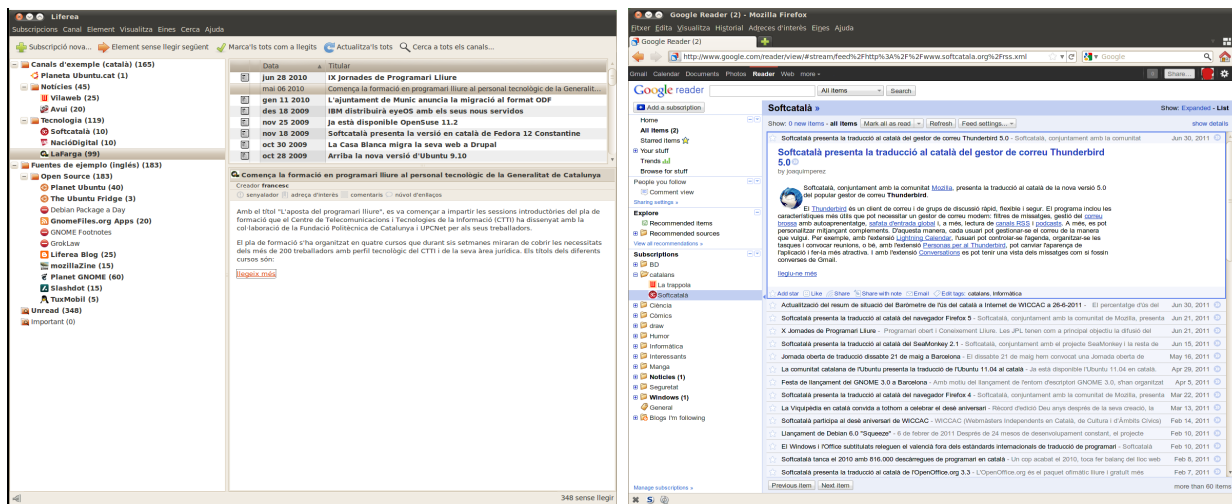


2. Centralizados *en servicios web o Agregadores web*. Hay que registrarse para poder acceder a las mismas funcionalidades que tienen los programas que se instalan. Tienen la ventaja de que podemos consultar nuestras sindicaciones desde cualquier ordenador con acceso a Internet. Algunos son:
 - www.google.com/reader
 - My Yahoo!
 - IGoogle
 - www.bloglines.com
 - www.netvibes.com
3. Integrados en el navegador web. Es una funcionalidad añadida e incorporada automáticamente por algunos navegadores:
 - Mozilla Firefox.
 - Opera.

A pesar de la división, normalmente el aspecto visual de estos programas es relativamente parecido. Todos suelen tener la pantalla dividida en bloques:

- En uno de los bloques suele haber la lista de suscripciones en la cual podemos agrupar estas suscripciones por temas.
- Un bloque para mostrar el contenido de cada una de las suscripciones.

En la imagen siguiente podemos ver el parecido entre el aspecto de un programa web como el Google Reader y un programa de escritorio como Liferea.



La gran diferencia suele estar en los filtros o las características extra que ofrecen cada uno de los programas para hacer el uso más interesantes a los usuarios: visualizaciones originales, estadísticas, etc.

Y dado que el volumen de información disponible es tan grande, resulta muy interesante el uso de algún tipo de catálogo que permita clasificar y realizar búsquedas sobre los contenidos de distintos feeds, necesitamos unas herramientas llamadas **directorios de canales**.

Estas dos herramientas pueden ser tanto online como herramientas para escritorio.

Un **blog** es una herramienta que recopila contenido ordenado cronológicamente.

Un **blogroll** es una colección de enlaces a blogs que, generalmente, aparecen listados en uno de los lados de la página web que muestra el blogroll.



ANEXO 1: Ejemplos de ficheros RSS y ATOM con ejemplos en distintas versiones.

DOCUMENTO RSS 0.91

RSS 0.91 es un dialecto de XML. Un documento RSS 0.91 permite presentar un único canal que puede contener hasta 15 ítems.

Cada ítem, a su vez, puede contener otras características como el título, la descripción, el lenguaje, etc.

Un ejemplo de DTD que recoge la parte más importante de RSS 0.91 es el siguiente:

```
<!ELEMENT rss (channel)>
<!ATTLIST rss version CDATA #REQUIRED>
<!ELEMENT channel (title | description | link | language
    | item+ | rating? | image?
    | textinput? | copyright?
    | pubdate? | lastBuidDate?
    | docs? | managingEditor?
    | webMaster? | skipHours?
    | skipDays?)*>
<!ELEMENT title (#PCDATA)>
<!ELEMENT description (#PCDATA)>
<!ELEMENT link (#PCDATA)>
<!ELEMENT image (title | url | link | width? | heigth? | description?)*>
<!ELEMENT url (#PCDATA)>
<!ELEMENT image (title | link | description)*>
<!ELEMENT name (#PCDATA)>
<!ELEMENT rating (#PCDATA)>
<!ELEMENT language (#PCDATA)>
<!ELEMENT width (#PCDATA)>
<!ELEMENT height (#PCDATA)>
<!ELEMENT copyright (#PCDATA)>
<!ELEMENT pubDate (#PCDATA)>
<!ELEMENT lastBuidDate (#PCDATA)>
<!ELEMENT docs (#PCDATA)>
<!ELEMENT managingEditor (#PCDATA)>
<!ELEMENT webMaster (#PCDATA)>
<!ELEMENT hour (#PCDATA)>
<!ELEMENT day (#PCDATA)>
<!ELEMENT skipHours (#PCDATA)>
<!ELEMENT skipDays (#PCDATA)>
```

En RSS 0.91 es obligatorio definir un **canal**.



Y hay 5 elementos **obligatorios** a la hora de definir un canal:

- title: Título.
- link: URL del contenido completo.
- description
- language
- image

Y otros **opcionales**, como copyright o rating.

Una vez definido el canal, se pueden definir elementos o ítems. En RSS 0.91 se pueden definir hasta 15 ítems.

Los ítems son los que definen el contenido que se quiere syndicar mediante los subelementos obligatorios (title, link, description).

Siguiendo el DTD anterior, un ejemplo de feed RSS 0.91 del IES El Caminàs es el siguiente:

```
<?xml version="1.0" encoding="UTF-8" ?>
<rss version="0.91">
  <channel>
    <title>IES EL CAMINAS</title>
    <link>http://www.ieselcaminas.org</link>
    <description>Novedades web del IES EL CAMINAS</description>
    <language>es-es</language>
    <copyright></copyright>
    <managingEditor>info@ieselcaminas.org</managingEditor>
    <webMaster>webmaster@ieselcaminas.org</webMaster>
    <pubDate>Mon, 19 Mar 12 09:00:00 GMT</pubDate>
    <image>
      <title>IES EL CAMINAS</title>
      <url>http://www.ieselcaminas.org/logotipo-caminas.png</url>
      <link>http://www.ieselcaminas.org </link>
      <width>80</width>
      <height>75</height>
      <description>Matriculas IES y Ciclos FProf</description>
    </image>
    <item>
      <title>Enseñanza Secundaria</title>
      <link>http://www.ieselcaminas.org/secundaria</link>
      <description>Las fechas de Matricula en Secundaria... </description>
    </item>
    <item>
      <title>Formación Profesional</title>
      <link>http://www.ieselcaminas.org/formacionprofesional</link>
      <description>Las fechas de Matricula en F. Profesional...</description>
    </item>
  </channel>
</rss>
```



```
</channel>
</rss>
```

DOCUMENTO RSS 0.92

El estándar RSS 0.92 es muy similar al RSS 0.91, pero presenta algunas diferencias:

- Permite representar un único canal, pero puede manejar un **número ilimitado de ítems** y soporta unos metadatos más completos y a nivel de ítem.
- Elimina la restricción de RSS 0.91 que limitaba el número de caracteres que pueden contener los elementos.
- Algunos elementos se convierten en opcionales, y aparecen otros nuevos para enriquecer el contenido:
 - **source**: para relacionar ítems con otros ítems de los que derivan.
 - **enclosure**: para definir un archivo asociado a un ítem.
 - **category**: para definir jerarquías de categorías. La jerarquía se representa separando las categorías con una barra invertida (p.ej. Categoría1/Categoría1.1/Categoría1.1.1).
 - **cloud**: sirve para definir los procesos de publicación y suscripción.

ELEMENTOS DE UN DOCUMENTO RSS 1.0

El mismo ejemplo anterior en RSS 1.0, es el siguiente:

```
<?xml version="1.0" encoding="UTF-8" ?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns="http://purl.org/rss/1.0"
xmlns:dc="http://purl.org/dc/elements/1.1"/>
  <channel rdf:about="http://www.ieselcaminas.org/caminas.rss">
    <title>IES EL CAMINAS</title>
    <link>http://www.ieselcaminas.org</link>
    <description>Novedades web del IES EL CAMINAS</description>
    <items>
      <rdf:Seq>
        <rdf:li resource="http://ieselcaminas.org/Secundaria"/>
        <rdf:li resource="http://ieselcaminas.org/FormacionProfesional"/>
      </rdf:Seq>
    </items>
  </channel>
  <item rdf:about="http://ieselcaminas.org/Secundaria"/>
    <title>Novedades fechas Matriculacion Secundaria</title>
    <link>http://www.ieselcaminas.org/Secundaria</link>
    <description>Las fechas de Matriculacion en Secundaria...</description>
```



```
<dc:date>2012-19-03</dc:date>
</item>
<item rdf:about="http://ieselcaminas.org/FormacionProfesional"/>
  <title>Novedades fechas Matriculacion Formación Profesional</title>
  <link>http://www.ieselcaminas.org/formacionprofesional</link>
  <description>Las fechas de Matriculacion en F.Profesional...</description>
  <dc:date>2012-19-03</dc:date>
</item>
</rdf:RDF>
```

En la primera etiqueta se definen los espacios que hacen que el documento sea más complejo pero a la vez más flexible y versátil.

Un feed RSS 1.0 debe comenzar con la línea:

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
```

Y contener al menos, un canal, que se define usando:

```
<channel rdf:about="URI_IDENTIFICADOR">
ELEMENTOS
</channel>
```

Donde URI_IDENTIFICADOR es un URI que sirve de identificador único para dicho canal.

RSS 1.0 soporta el uso de módulos para una mayor flexibilidad. Un ejemplo de módulo es el siguiente: módulo **mod_annotation** que permite añadir un nuevo elemento donde se puede definir una URL en la que se puede discutir sobre el contenido del ítem al que acompaña. Para usarlo hay que añadir el espacio de nombres correspondiente:

```
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns="http://purl.org/rss/1.0"
xmlns:annotate="http://purl.org/rss/1.0/modules/annotate/">
```

y se podrá añadir el elemento siguiente:

```
<annotate:reference rdf:resource="URL_DISCUSION"/>
```

ELEMENTOS DE UN DOCUMENTO RSS 2.0

Sigue la línea de RSS 0.91 pero tiene las siguientes diferencias:



- Su estructura está pensada para proporcionar una mayor modularidad, y lo consigue con la introducción de módulos.
Esto provoca que los documentos que genera este estándar sean más complejos.
- Incorpora un conjunto de nuevos elementos, *comments*, *author* y *guid*, entre otros. Por ej. *guid* se recomienda usarlo ya que permite definir una URL única para el ítem al que acompaña mejorando la interoperabilidad entre sistemas.
- Y en los elementos que ya existían, se introducen algunos cambios: por ejemplo, el elemento *image* ahora es opcional.

Como sigue la línea de RSS 0.91 y RSS 0.92, al RSS 2.0, se le ha llamado, a veces, RSS 0.94 (RSS 0.93 fue un desarrollo que se abandonó al poco tiempo de aparecer)

La inclusión de módulos no es como en RSS 1.0, que usaba RDF. En RSS 2.0 no se usa RDF. Pero sí que coinciden en que la definición del espacio de nombres correspondiente se incluye en la declaración del elemento raíz.

El primer módulo que se apareció en el estándar RSS 2.0 fue *blogChannel* para permitir incluir información en el *feed* acerca de blogs. Para poder usarlo hay que incluir su espacio de nombres en el elemento raíz de la siguiente forma:

```
<rss version="2.0" xmlns:blogChannel="http://backend.userland.com/blogChannelModule">
```

Una vez incluido el espacio de nombres, se pueden usar tres nuevos elementos:

- *blogChannel:blogRoll*: contiene una URL que apunta al *blogroll* del sitio
- *blogChannel:blink*: contiene una URL a un sitio web que el autor considera interesante.
- *blogChannel:mySubscriptions*: contiene una URL al fichero que contiene los feeds RSS a los que el autor del sitio está suscrito.

Siguiendo con el mismo ejemplo anterior, la versión RSS 2.0 del feed del IES EL CAMINAS:

```
<?xml version="1.0" encoding="UTF-8" ?>
<rss version="2.0" xmlns:blogChannel="http://backend.userland.com/blogChannelModule">
  <channel>
    <title>IES EL CAMINAS</title>
    <link>http://www.ieselcaminas.org</link>
    <description>Novedades web del IES EL CAMINAS</description>
    <blogChannel:blogRoll>
      http://www.ieselcaminas.org/blogroll.opml
    </blogChannel:blogRoll>
    <blogChannel:blink>
      http://www.edu.gva.es/
    </blogChannel:blink>
    <blogChannel:mySubscriptions>
      http://www.ieselcaminas.org/subscriptions.opml
    </blogChannel:mySubscriptions>
  </channel>
</rss>
```



```

</blogChannel:mySubscriptions>
<language>es-es</language>
<copyright></copyright>
<managingEditor>info@ieselcaminas.org</managingEditor>
<webMaster>webmaster@ieselcaminas.org</webMaster>
<pubDate>Mon, 19 Mar 12 09:00:00 GMT</pubDate>
<image>
  <title>IES EL CAMINAS</title>
  <url>http://www.ieselcaminas.org/logotipo-caminas.png</url>
  <link>http://www.ieselcaminas.org </link>
  <width>80</width>
  <height>75</height>
  <description>Matriculas IES y Ciclos FProf</description>
</image>
<item>
  <title>Enseñanza Secundaria</title>
  <link>http://www.ieselcaminas.org/secundaria</link>
  <description>Las fechas de Matricula en Secundaria... </description>
  <author>info@ieselcaminas.org</author>
</item>
<item>
  <title>Formación Profesional</title>
  <link>http://www.ieselcaminas.org/formacionprofesional</link>
  <description>Las fechas de Matricula en F. Profesional...</description>
</item>
</channel>
</rss>

```

DOCUMENTO ATOM

Un ejemplo de fichero ATOM 1.0 de sindicación puede ser el siguiente:

```

<?xml version="1.0" encoding="UTF-8"?>
<feed xmlns="http://purl.org/atom/ns#draft-ietf-atompub-format-07">

  <title>Example Feed</title>
  <link href="http://example.org/" />
  <updated>2003-12-13T18:30:02Z</updated>

```



```
<author>
  <name>John Doe</name>
</author>

<entry>
  <title>Atom-Powered Robots Run Amok</title>
  <link href="http://example.org/2003/12/13/atom03"/>
  <id>urn:uuid:1225c695-cfb8-4ebb-aaaa-80da344efa6a</id>
  <updated>2003-12-13T18:30:02Z</updated>
  <content>Some text.</content>
</entry>

</feed>
```



ANEXO 2: Evolución de las distintas versiones

---1997	Antes de RSS, existían otros formatos (o lenguajes) que soportaban la sindicación: CDF. Utilizando CDF podíamos suscribirnos a varios sitios con publicaciones online. Internet Explorer comprobaba estas publicaciones de forma regular y descargaba los nuevos datos a medida que estaban disponibles.
1997	Dave Winer , desarrollador de Userland, introdujo su propio formato de sindicación basado en XML en su blog de noticias Scripting New.
1999	La primera versión verdadera de RSS fue creada por Dan Libby de Netscape. Se convirtió en la versión 0.90 (diseñada para emplearse en el portal de My Netscape).
1999	En julio, la versión 0.90 se modificó dando lugar a la aparición de la primera versión popular, la 0.91 . En ese momento, RSS se conocía con el nombre de Rich Site Summary (Compendio enriquecido de sitios). La versión 0.91 adoptó parte del formato de Dave Winer y se convirtió en una de las principales versiones de RSS.
1999	Netscape abandonó RSS y apareció una lista de correo llamada RSS-DEV que se hizo muy popular. Al mismo tiempo, Dave Winer creó una versión modificada de RSS 0.91 , que ya se había utilizado en el software de Userland.
2000	El grupo RSS-DEV creó la versión 1.0 de RSS. Esta versión se alejaba bastante de las versiones anteriores y basada en el lenguaje RDF (Formato de descripción de recursos). RSS se interpretaron como RDF Site Summary (Compendio RDF de sitios). Pero era muy complicado y no fue bien aceptado.
2000	Sólo tres semanas después del anuncio de RSS 1.0, Dave Winer lanzó la versión 0.92 de RSS. RSS 1.0 fue también una versión importante, pero su sintaxis fue tan diferente de las otras versiones de RSS que la gente la encontró demasiado difícil de utilizar.
2001	En abril, Dave Winer lanzó RSS 0.93 , similar a la versión 0.92.
2001	En agosto, apareció un borrador de la versión 0.94 , que eliminaba los cambios realizados en la versión 0.93 e introducía unos cuantos cambios menores.
2002	Dave Winer lanzó un sucesor de la versión 0.92 conocida como versión 2.0 (el nombre de versión 1.0 estaba ya cogido). En ese momento, RSS comenzó a conocerse como Real y Simple Syndication (Sindicación realmente sencilla).
2003	Dave Winer transmitió la propiedad de RSS 2.0 al Berkman Center for Internet & Society de la Harvard Law School, donde trabajaba en ese momento. RSS 2.0 se estabilizó desde entonces.
2005	Pero siguió la evolución de RSS. Palmer y Schmidt lanzaron un borrador de RSS 1.1, que simplificaba el lenguaje en cierta medida. En noviembre de 2005, Microsoft propuso algunas extensiones de RSS, llamadas de forma informal Real Simple Synchronization (Sincronización sencilla real).
2006	Varios desarrolladores lanzaron un nuevo formato que se llamó Atom como sustituto del formato RSS. Atom está basado en XML, pero es bastante más difícil de aprender y es más complejo que RSS. Atom mejora RSS añadiendo capacidades de internacionalización, estandarizando la sintaxis y permitiendo que los usuarios añadan sus propias características. Sin embargo, Atom todavía tiene que realizar un largo camino, en comparación con RSS.
2012	El desarrollo de RSS sigue en marcha. Hoy en día, las versiones 0.91, 1.0 y 2.0, así como Atom se utilizan mucho.