LMSGI

Bloque 3 - Aplicación de los lenguajes de marcas a la sindicación de contenidos











"Una manera de hacer Europa". Cofinanciación a cargo del Programa Operativo del FSE 2014-2020 para Extremadura gastos de Ciclos Formativos de Grados Medio y Superior.



Índice





- Introducción. Ventajas y ámbitos de aplicación
- Estructura de los canales de contenidos
- Tecnologías de creación de canales de contenidos
- Sindicación en tres pasos
- Etiqueta <link> en página Web y sindicación
- <![CDATA[....]]>
- Herramientas online para generación canales/feed desde página Web
- Validación RSS y ATOM
- Directorios de canales de contenidos
- Agregador
- Sitemap y relación con RSS









Introducción. Ventajas y ámbitos de aplicación

- Sindicación web (redifusión) permite a un sitio Web ofrecer sus contenidos a otros (textos, audios, vídeos, ..). El lector puede ser otra web o un software específico, denominado agregador. El sitio web o lector receptor recibe un resumen (web feeds) de las últimas informaciones del sitio emisor. Esta facilidad permite a usuarios de Internet consumir en un solo sitio y de manera resumida solamente el contenido de su interés de todos los sitios web que visite. [Fuente wiki] Más información
- La redifusión suele realizarse bajo una licencia o contrato entre las partes.
- Las fuentes suelen codificarse en lenguaje XML, lo habitual es utilizar RSS o ATOM
- Muy utilizado en blogs y en web de noticias, por ejemplo, <u>El País</u>, <u>Hoy</u>, ...
- Ayuda a reutilizar contenidos, a aumentar las visitas, a establecer relaciones con otros sitios, a mejorar el posicionamiento de la web en los buscadores,
- Artículo en inglés sobre la vigencia de la sindicación de contenidos.
- OPML (Outline Processor Markup Language) → listar varias fuentes juntas, generador <u>https://opml-gen.ovh/</u> ← *Trabajo voluntario*







```
<?xml version="1.0"?>
<!-- RSS generated by Radio UserLand v8.0.5 on 9/30/2002; 4:00:00 AM Pacific -->
<rss version="2.0" xmlns:blogChannel="http://backend.userland.com/blogChannelModule">
                                                                                                                                               <?xml version="1.0" encoding="utf-8"?>
                                                                                                                                               <rss version="2.0">
       <title>Scripting News</title>
       k>http://www.scripting.com/</link>
                                                                                                                                                 <channel>
       <description>A weblog about scripting and stuff like that.</description>
                                                                                                                                                    <!-- Aquí van las propiedades del canal -->
       <language>en-us</language>
       <blogChannel:blogRoll>http://radio.weblogs.com/0001015/userland/scriptingNewsLeftLinks.opml</blogChannel:blogRoll>
       <!-- Aquí va el contenido de una entrada publicada -->
       <blogChannel:blink>http://diveintomark.org/</blogChannel:blink>
       <copyright>Copyright 1997-2002 Dave Winer</copyright>
                                                                                                                                                    </item>
       <lastBuildDate>Mon, 30 Sep 2002 11:00:00 GMT</lastBuildDate>
       <docs>http://backend.userland.com/rss</docs>
                                                                                                                                                    ... <!-- Tantos elementos como entradas queramos -->
       <generator>Radio UserLand v8.0.5</generator>
                                                                                                                                                    <item>
       <category domain="Syndic8">1765</category>
       <managingEditor>dave@userland.com</managingEditor>
                                                                                                                                                    </item>
       <webMaster>daye@userland.com</webMaster>
                                                                                                                                                 </channel>
       <++1>40</++1>
       <item>
           <description>&quot;rssflowersaliqnriqht&quot;With any luck we should have one or two more days of namespaces stuff here on Scripting News
it's winding down. Later in the week I'm going to a alt; a href=aguot; http://harvardbusinessonline.hbsp.harvard.edu/b02/en/conferences/conf detail.jhtml?
id=s775stg&pid=144XCF">conference</a&gt; put on by the Harvard Business School. So that should change the topic a bit. The following week I'm off to
Colorado for the &lt:a href=&guot:http://www.digitalidworld.com/conference/2002/index.php&guot:&gt:Digital ID World&lt:/a&gt: conference. We had to go through
namespaces, and it turns out that weblogs are a great way to work around mail lists that are clogged with <a
href="http://www.userland.com/whatIsStopEnergy"&qt;stop energy</a&qt;. I think we solved the problem, have reached a consensus, and will be ready to move
forward shortly.</description>
           <pubDate>Mon, 30 Sep 2002 01:56:02 GMT</pubDate>
           <quid>http://scriptingnews.userland.com/backissues/2002/09/29#When:6:56:02PM</quid>
           </item>
       <item>
           <description>Joshua Allen: &lt:a href=&guot:http://www.netcrucible.com/blog/2002/09/29.html#a243&guot:&gt:Who loves namespaces?&lt:/a&gt:</description>
           <pubDate>Sun, 29 Sep 2002 19:59:01 GMT</pubDate>
           <quid>http://scriptingnews.userland.com/backissues/2002/09/29#When:12:59:01PM</quid>
           </item>
       <item>
           <description>&lt;a href=&quot;http://www.docuverse.com/blog/donpark/2002/09/29.html#a68&quot;&qt;Don Park&lt;/a&qt;: &quot;It is too easy for engineer to
anticipate too much and XML Namespace is a frequent host of over-anticipation. " </description>
           <pubDate>Mon, 30 Sep 2002 01:52:02 GMT</pubDate>
           <guid>http://scriptingnews.userland.com/backissues/2002/09/29#When:6:52:02PM</guid>
```

RSS 2.0 en Ganbeta

Declaración del documento XML y de la especificación a utilizar (RSS 2.0, ATOM 1.0, ...) + Channel/Feed + Item-s/Entry-s

</item>

Tecnologías de creación de canales de contenidos: RSS y ATOM





- Todavía es más utilizado RSS (de hecho se suele hablar de RSS aunque se utilice ATOM) aunque <u>ATOM</u> está creciendo y técnicamente es preferible, por ejemplo, es la opción utilizada por Google.
- Es común que la fuente ofrezca sus contenidos en RSS y en ATOM, por ejemplo:
 - https://www.hoy.es/rss/
- La mayoría de los sitios web están realizados con sistemas tipo <u>CMS</u> o similares, por ejemplo, Wordpress o Joomla. Estos sistemas permiten crear y administrar canales de sindicación a los que se van añadiendo de forma automática todos los contenidos nuevos. Nota: ASIR y DAW trabaja en 2º con CMS
- Directrices para feeds (Google)
- Más información
- Comparación RSS ATOM

RSS 2.0	Atom 1.0	
author	author *	
category	category	
channel	feed	
copyright	rights	
-	subtitle	
description *	summary and/or content	
generator	generator	
guid	id*	
image	logo	
item	entry	
lastBuildDate (in channel)	updated *	
link *	link *	
managingEditor	author or contributor	
pubDate	published (subelement of entry)	
title *	title *	
ttl	_	

PORTADA 🔯 🔯			
ÚLTIMA HORA 🔯 🛭	1		
40 Aniversario de la	Constitución 🔝 🔊		
40 Aniversario de la Agro 🔯 🔯	Constitución M M		
	Constitución 🔝 🔊	Análisis	<u> </u>
Agro 🔝 🔯		Análisis Ganadenia	a 8
Agro 🔯 🔯	<u>a</u> a		

* elementos obligatorios

Tecnologías de creación de canales de contenidos - RSS (I)





- RSS Historia → no sigue un proceso de estandarización reconocido, genera confusión.
- Rich Site Summary (RSS 0.91) (Resumen óptimo del sitio)
- RDF Site Summary (RSS 0.9 y 1.0) (Resumen RDF del sitio)
- Really Simple Syndication (<u>RSS 2.0 (</u>Sindicación realmente simple) → **es la que utilizaremos**
- RSS utiliza XML y contiene datos de la novedades del sitio, como el título, fecha de publicación o la descripción.
- El elemento **<channel>** al menos tiene que incluir: <title>, , <description> y 1 o más <item>
- Cada <item> al menos: <title>, <link> y <description>
- Es recomendable que incluya la fecha de publicación: <pubDate>
- Documentación RSS 2.0:
 - Especificación RSS 2.0 Ejemplo fichero RSS 2.0 RSS 2.0 Google RSS 2.0 IONOS RSS en
 W3Schools Resumen RSS 2.0 RSS Best Practices



Version	Owner	Pros	Status	Recommendation
0.90	Netscape		Obsoleted by 1.0	Don't use
0.91	UserLand	Drop dead simple	Officially obsoleted by 2.0, but still quite popular	Use for basic syndication. Easy migration path to 2.0 if you need more flexibility
0.92, 0.93, 0.94	UserLand	Allows richer metadata than 0.91	Obsoleted by 2.0	Use 2.0 instead
1.0	RSS-DEV Working Group	RDF-based, extensibility via modules, not controlled by a single vendor	Stable core, active module development	Use for RDF-based applications or if you need advanced RDF-specific modules
2.0	UserLand	Extensibility via modules, easy migration path from 0.9x branch	Stable core, active module development	Use for general-purpose, metadata-rich syndication



Tecnologías de creación de canales de contenidos - RSS (y II)





https://www.w3schools.com/xml/xml_rss.asp#rssref

<u>http://www.rssboard.org/rss-profile#element-channel-item-link</u> → Podemos ver los que son obligatorios y opcionales, por ejemplo, en item, description y title solo es obligatorio uno de los dos, y link no es obligatorio. **URL tienen que ser absolutas**

RSS <channel> Element The links in the "Element" column point to more information about each specific element.</channel>			
Element	Description		
<category></category>	Optional. Defines one or more categories for the feed		
<cloud></cloud>	Optional. Register processes to be notified immediately of updates of the feed		
<copyright></copyright>	Optional. Notifies about copyrighted material		
<description></description>	Required. Describes the channel		
<docs></docs>	Optional. Specifies a URL to the documentation of the format used in the feed		
<generator></generator>	Optional. Specifies the program used to generate the feed		
<image/>	Optional. Allows an image to be displayed when aggregators present a feed		
<language></language>	Optional. Specifies the language the feed is written in		
< <u>lastBuildDate></u>	Optional. Defines the last-modified date of the content of the feed		
k>	Required. Defines the hyperlink to the channel		
<managingeditor></managingeditor>	Optional. Defines the e-mail address to the editor of the content of the feed		
<pubdate></pubdate>	Optional. Defines the last publication date for the content of the feed		
<rating></rating>	Optional. The PICS rating of the feed		
<skipdays></skipdays>	Optional. Specifies the days where aggregators should skip updating the feed		
<skiphours></skiphours>	Optional. Specifies the hours where aggregators should skip updating the feed		
<textinput></textinput>	Optional. Specifies a text input field that should be displayed with the feed		
<title></td><td>Required. Defines the title of the channel</td></tr><tr><td><tti></td><td>Optional. Specifies the number of minutes the feed can stay cached before refreshing it from the source</td></tr><tr><td><webMaster></td><td>Optional. Defines the e-mail address to the webmaster of the feed</td></tr></tbody></table></title>			

lement	Description
author>	Optional. Specifies the e-mail address to the author of the item
category>	Optional. Defines one or more categories the item belongs to
comments>	Optional. Allows an item to link to comments about that item
description>	Required. Describes the item
enclosure>	Optional. Allows a media file to be included with the item
g <u>uid></u>	Optional. Defines a unique identifier for the item
<u> </u>	Required. Defines the hyperlink to the item
pubDate>	Optional. Defines the last-publication date for the item
source>	Optional. Specifies a third-party source for the item
title>	Required, Defines the title of the item

Formato de fechas en RFC 822

<pubDate>Mon, 15 Oct 2007 14:10:00 GMT</pubDate>

<lastBuildDate>Mon, 15 Oct 2007 09:10:00 EST/lastBuildDate>

<pubDate>Mon, 15 Oct 2007 08:10:00 -0600</pubDate>

Fecha en diferentes formatos

https://www.utctime.net/ y Epoch Unix Timestamp Converter Online

Tecnologías de creación de canales de contenidos - ATOM





- Especificación elaborada por IETF
- También utiliza XML
- Documentación:
 - o RFC 4287 Especificación
 - ATOM 1.0 en Google
 - Documentación en W3C: <u>Introduction to ATOM</u> y <u>The ATOM Syndication Format</u>
 - <u>Directriz Junta de Andalucía para el uso de ATOM</u>
- <feed>
 - Obligatorio: id, title, updated
 - **Recomendado:** author y <u>link</u> (ruta absoluta)
- <entry>
 - Obligatorio: id, title, updated
 - **Recomendado:** author, content, <u>link</u> y summary
- Formato de Fechas RFC-3339:

```
Example Date constructs:

<updated>2003-12-13T18:30:02Z</updated>
<updated>2003-12-13T18:30:02.25Z</updated>
<updated>2003-12-13T18:30:02+01:00</updated>
<updated>2003-12-13T18:30:02.25+01:00</updated>
```

```
<?xml version="1.0" encoding="utf-8"?>
<feed xmlns="http://www.w3.org/2005/Atom">
   <title>Example Feed</title>
   <subtitle>A subtitle.</subtitle>
   k href="http://example.org/feed/" rel="self" />
   k 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>
        k href="http://example.org/2003/12/13/atom03" />
       k rel="alternate" type="text/html"
href="http://example.org/2003/12/13/atom03.html"/>
       k rel="edit" href="http://example.org/2003/12/13/atom03/edit"/>
        <id>urn:uuid:1225c695-cfb8-4ebb-aaaa-80da344efa6a</id>
        <updated>2003-12-13T18:30:02Z</updated>
        <summary>Some text.</summary>
        <content type="xhtml">
           <div xmlns="http://www.w3.org/1999/xhtml">
               This is the entry content.
           </div>
       </content>
        <author>
           <name>John Doe</name>
           <email>johndoe@example.com</email>
       </author>
   </entry>
```

Fecha en diferentes formatos https://www.utctime.net/ y Epoch Unix Timestamp Converter Online

Sindicación en tres pasos





- 1. Creación de un canal (feed)
- 2. Añadir < link > en la página-s HTML (ver diapositiva siguiente)
- 3. Añadir un hipervínculo <a> a la página-s HTML → el enlace que ponemos en link> solo se ve si editamos el código o con software de agregación, por ello conviene añadir un enlace en la propia página, por ejemplo, en RSS sería:

```
<a rel="alternate" type="application/rss+xml" href="http://example.com/rss"> ...</a>
<a rel="alternate" type="application/rss+atom" href="http://example.com/feed">...</a>
```

<u>http://feedicons.com/</u> → iconos en diferentes colores para indicar un enlace a sindicación





Etiqueta < link > en página Web y sindicación

- Las páginas Web que ofrecen sus contenidos por sindicación, utilizan la etiqueta:

```
<IDOCTYPE html>
<html lang="es">
<head>
<meta charset="UTF-8">
<meta name="viewport" content="width=device-width, initial-scale=1">
<link rel="profile" href="http://gmpg.org/xfn/11">
<link rel="profile" href="http://informatica.iesvalledeljerteplasencia.es/wordpress/xmlrpc.php">
<title>Ciclos Formativos IES Valle del Jerte &#8211; Plasencia &#8211; Contenidos elaborados por alumnos</title>
<link rel='dns-prefetch' href='//fonts.googleapis.com' />
<link rel='dns-prefetch' href='//s.w.org' />
<link rel="alternate" type="application/rss+xml" title="Ciclos Formativos IES Valle del Jerte - Plasencia &raquo; Feed"
href="http://informatica.iesvalledeljerteplasencia.es/wordpress/feed/" />
```







En un documento XML, una sección CDATA evita que el texto que contiene sea interpretado por el analizador sintáctico. No hay diferencia semántica entre una cadena de caracteres dentro de una sección CDATA y la sintaxis usual en la que "<" y "&" estarían representados por "& It;" y "& amp;", respectivamente. Fuente: https://es.wikipedia.org/wiki/CDATA



https://www.w3.org/TR/REC-xml/ #sec-cdata-sect

▼ <description>

</description>

<pubDate>2019-10-15T09:45:08+02:00 </pubDate>

El servicio del AVE en Girona sique interrumpido por daños en la infraestructura

Herramientas online para generación canales/feed desde página Web

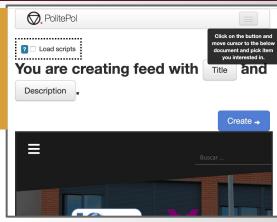






PolitePol nos permite de forma sencilla crear un canal de una página web.

Editor código RSS Online



https://www.mysitemapgenerator.com/rss/create.html





```
version="1.0" encoding="utf-8"?>
version="2.0" xmlns:content="http://purl.org/rss/1.0/modules/content/" xmlns:dc="http://purl.org/dc/elements/1.1/">
       le>Impresión de páginas web Ciclos Formativos IES Valle del Jerte Plasencia</title>
|>>! [CDATA[http://informatica.iesvalledeljerteplasencia.es]]>>/link>
              n>Impresión de páginas web Ciclos Formativos IES Valle del Jerte Plasencia</description>
          e>Tue, 01 Oct 2019 13:30:49 +0300</pubDate>
             >MySitemapGenerator (www.mysitemapgenerator.com)</generator>
            isPermaLink="false">173437457</quid>
            e>Ciclos Formativos IES Valle del Jerte Plasencia</title>
            >http://informatica.iesvalledeljerteplasencia.es/wordpress/que-es-el-seo-y-por-que-lo-necesito/</link>
                  >Administrator</dc:c
                >Site News</ca
               >Tue, 01 Oct 2019 13:30:51 +0300</pubDate>
             isPermaLink="false">446196940</gr
            e>Impresión de páginas web Ciclos Formativos IES Valle del Jerte Plasencia</title>
            >http://informatica.iesvalledeljerteplasencia.es/wordpress/impresion-de-paginas-web/</link>
                  r>Administrator</dc:c
               >Tue, 01 Oct 2019 13:30:52 +0300</pubDate>
             isPermaLink="false">370462470</guid>
             >Ciclos Formativos IES Valle del Jerte Plasencia</title>
            >http://informatica.iesvalledeljerteplasencia.es/wordpress/normas-de-estilo/</link>
                  r>Administrator</dc:creat
```







https://validator.w3.org/feed/

Feed Validation Check the syntax of Atom or		
Validate by URI Validate by URI Validate a feed online: Address:	ut .)://www.w3.org/2005/Atom">)://www.w3.org/2005/Atom">):xt">dive into mark ="html"> .ot of effort cing this effortless)7-31T12:29:29Z
,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	Check	:.org,2003:3 :.org,2003:3 :rnate" type="text/html" href="http://example.org/"/> F" type="application/atom+xml" example.org/feed.atom"/> ht (c) 2003, Mark Pilgrim ="http://www.example.com/"
This is the W3C Feed Validation Service, a f regular Web pages.	ree service that checks the syntax of Atom or RSS feeds. The Markup Validation Service is also available if you wish to validate	"http://www.example.com/"
I.♥ VALIDATOR	The W3C validators rely on community support for hosting and development. Donate and help us build better tools for a better web.	
	Atom This is a valid Atom 1.0 feed.	

Recommendations

LMSGI - Antonio Berrocal Piris

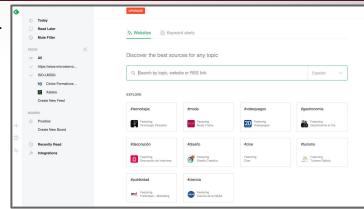
This feed is valid, but interoperability with the widest range of feed readers could be improved by implementing the following recommendations. line 13, column 40: Self reference doesn't match document location [help]

Directorios de canales de contenidos





- Permite dar a conocer los canales, clasificándolos por categorías.
- Algunos agregadores, por ejemplo, https://feedly.com/, incluyen buscadores que nos ofrecen listados de canales/feeds a los que nos podemos suscribir.
- También podemos encontrar listados de feeds/canales por ejemplo: https://www.specificfeeds.com/, https://www.ontoplist.com/, https://www.ontoplist.com/, https://www.wingee.com/, https://www.winge
- En otras web, como <u>https://www.toprankblog.com/rss-blog-directories/</u>, podemos ver listados de directorios
- Es habitual intercambiar la dirección de nuestro canal con el de otros, de forma que en nuestra web incluimos un enlace a otro canal, a cambio en la web del canal incluyen un enlace al nuestro.



Aparecer en directorios o en otras web ayuda a mejorar la posición de nuestra web en los buscadores (SEO, Search Engine Optimization).

Más información en https://feeder.co/knowledge-base/rss-basics/what-is-an-rss-directory/

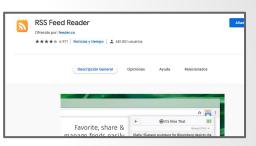
Agregador

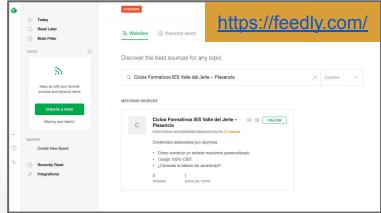




- Agregador: aplicación web o de escritorio que muestra los contenidos "nuevos" publicados por los emisores de sindicación de contenidos a los que estamos suscritos.
- Podemos encontrar agregadores online, extensiones para los navegadores o versiones de escritorio. Más información en Alternativas a Digg Reader: los mejores lectores de feeds y aplicaciones RSS de 2018
- El agregador comprueba periódicamente las actualizaciones que hay en los canales a los que estamos suscritos.
- La mayoría de los navegadores incluían un agregador, pero las versiones actuales lo han suprimido para simplificar el mantenimiento del software. Nota: noticia <u>Firefox</u> → podemos utilizar una extensión







Sitemap y relación con RSS





- Fuente: de sitemaps.org y Crear y enviar un sitemap | Centro de la Búsqueda de Google
- Los sitemaps son archivos, normalmente XML, en los que se proporciona información sobre las páginas, los vídeos y otros archivos de un sitio web, así como sobre las relaciones entre ellos. Los buscadores, como Google, leen estos archivos para rastrear los sitios web de forma más eficaz. Los sitemaps informan a los rastreadores de qué archivos de un sitio web son importantes según el webmaster y, además, incluyen datos importantes sobre ellos; por ejemplo, en el caso de las páginas, pueden indicar cuándo se actualizaron por última vez, cada cuánto se modifican y si existen versiones en otros idiomas. Las entradas de vídeo pueden indicar la duración, la categoría y la clasificación de edad recomendada de los diferentes vídeos. Las entradas de imagen pueden incluir el tema, el tipo y la licencia de las distintas imágenes.
- Se necesita un sitemap si el sitio es muy grande, tiene muchas páginas no enlazadas con otras o es nuevo y hay pocos enlaces externos que lo enlacen. <u>Más información.</u>
- Lo recomendable es que utilice el <u>formato "estándar" XML sitemap</u>, pero se soportan otros formatos, en el caso de Google, <u>acepta como sitemaps RSS y ATOM</u>
- Si el índice es muy grande, en el caso de Google, tiene más de 50.000 URL, se tiene que dividir en varios y utilizar un <u>índice de sitemaps</u>
- Herramientas para crear sitemap o https://www.xml-sitemaps.com/
- Validador de sitemaps: propias herramientas del webmaster de Google u otros y en https://www.xml-sitemaps.com/validate-xml-sitemap.html
- Formato de fechas: <u>Date and Time Formats</u> de W3C

