

Tema 7. Sindicación de contenidos

1. Sindicación.....	3
1.1. Concepto de sindicación.....	3
1.2. Suscripción.....	4
1.3. Un ejemplo.....	4
2. Formatos.....	5
2.1. Formato RSS.....	5
2.2. Formato Atom.....	6
2.3. Validación.....	6
2.4. Editar feeds.....	6

Tema 7. Sindicación de contenidos.

1. Sindicación.

La velocidad a la que fluye la información en Internet, hace que muchas veces ésta se escape a nuestro control, o bien no cubra nuestras expectativas o que realicemos una navegación muy engorrosa, compleja o innecesaria. Un claro ejemplo es el de aquellas personas que suelen tener una serie de páginas favoritas que visitan frecuentemente en busca de novedades, sobre todo, páginas personales y blogs. En muchas ocasiones estas visitas son estériles porque no contiene nada nuevo desde la última vez que se visitaron.

Si en vez de visitar periódicamente dichos sitios web en busca de novedades, son éstos los que avisan a las personas cuando las novedades se producen, se habrá conseguido simplificar y disminuir los tiempos de navegación.

1.1. Concepto de sindicación.

La **Redifusión web** (o **sindicación web**) consiste en el reenvío o reemisión de contenidos desde una fuente original (sitio web de origen) hasta otro sitio web de destino (receptor) que, a su vez, se convierte en emisor puesto que pone a disposición de sus usuarios los contenidos a los que en un principio sólo podían tener acceso los usuarios del sitio web de origen.

El término *sindicación web* se emplea con mayor frecuencia que el de *redifusión web*, especialmente, en lo referido a contenidos web. De igual manera, la fuente origen se suele denominar *fuentes web* cuando el contenido informativo es una página web (páginas de noticias, blogs, comentarios de foros, etc.)

El contenido para la redifusión de las fuentes web suele codificarse en XML. Las dos principales familias de formatos de redifusión web son *RSS* y *Atom*, con sintaxis muy parecidas.

El formato *Atom* fue desarrollado como una alternativa a *RSS*, debido a que existían incompatibilidades entre algunas versiones de este último. Los que propusieron el formato *Atom* organizaron el grupo de trabajo *IETF Atom Publishing Format and Protocol*, y la recomendación se publicó como *RFC 5023*.

Un documento *Atom* puede contener mayores volúmenes información, incluso con mayores niveles de complejidad, superando siempre su consistencia frente a su documento *RSS* equivalente. Las mejoras que ofrece *Atom* respecto a cualquier versión de *RSS* hacen que su uso empezara a extenderse rápidamente a pesar de ser algo más complicado. Sin embargo, el formato *RSS* ha seguido siendo más popular y se ha utilizado más.

El lenguaje *RSS* consta de un vocabulario que permite describir información de tal manera que puede ser reutilizada. Actualmente, *RSS* es acrónimo de "*Really Simple Syndication*", aunque anteriormente lo fue de "*Rich Site Summary*" y de "*RDF Site Summary*". Estos cambios de denominación se deben a la aparición de distintos estándares:


- **Rich Site Summary** (*RSS 0.91, 0.92, 0.94*)
- **RDF Site Summary** (*RSS 0.9 y 1.0*) basado en la *Especificación RSS 1.0*: <http://purl.org/rss/1.0> que cumple con la especificación *RDF*.
- **Really Simple Syndication** (*RSS 2.0*) basado en la *Especificación RSS 2.0*. <http://blogs.law.harvard.edu/tech/rss> que no cumple con la especificación *RDF*.

Tanto *RSS* como *Atom* son un formato para distribuir un conjunto de titulares organizados en forma de sumario o índice, llamados canales (*channels* o *feeds*), a cuyos contenidos se puede acceder a través de Internet sin necesidad de usar el navegador. De esta forma, el usuario dispone de los titulares actualizados de muchos sitios webs sin la necesidad de conectarse a cada uno de ellos. Muchas webs están adaptadas a *RSS* y/o *Atom* permitiendo el acceso a una enorme cantidad de información de forma automatizada.

Un *canal* (*RSS*) o *feed* (*Atom*), siempre contiene el título de la cabecera, un breve resumen de la información del *canal* o *feed* y un enlace que conduce al contenido o texto completo de la noticia. Los enlaces a canales o *feeds* se suelen indicar en los sitios web utilizando iconos que contienen las siglas *RSS* o *Atom*.

Los documentos, al tener formato *XML*, son archivos de texto, albergados en el servidor del sitio web que ejerce de fuente de noticias. Dentro de estos archivos, los encargados de editar el sitio web (webmasters) colocan, de una forma ordenada y estructurada, los titulares y una pequeña descripción o entrada de los últimos contenidos que han publicado en su web.

Así pues, además de los titulares, también contienen una serie de códigos que facilitan a los programas que leen estos archivos la detección de nuevos contenidos. Esto permite avisar al usuario, de forma automática, de la disponibilidad de nuevas noticias.

Hay programas específicos para leer estos formatos, llamados *agregadores*. Un *agregador* se suscribe a una serie de *canales* o *feeds*, y te presenta exclusivamente la información contenida en dichos canales. Las direcciones de los *feeds* o *canales* vienen habitualmente indicados por iconos de este tipo , que muestran *XML*, *RSS* o *RDF*, como en este caso. Otros iconos permiten añadir directamente el *canal* o *feed* a tu *agregador* preferido.

El *agregador* o *lector de feed* podría ser un *lector RSS* instalado en el ordenador (un programa o software específico), un navegador web (como *Internet Explorer*), un programa de correo electrónico (como *Thunderbird*) o un lector online (una página web específica en la que se puede añadir varios canales y ver las noticias juntas).

Es importante destacar que cada vez que los artículos de un canal suscrito se actualizan, se reflejarán automáticamente en lector, sin necesidad de que el usuario realice ninguna operación adicional. Además, el lector alertará de dicha actualización de contenidos.

En definitiva, sindicación web es una forma de suscripción en la cual parte del material de un sitio web se hace disponible en otros sitios web o de un lector especial. 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, sólo el contenido que es de su interés de aquellos sitios web que visita.

1.2. Suscripcion.


Como lector de *feeds* podemos usar un explorador web pero, si queremos que se nos informe de forma automática cuando se produce una novedad o noticia en el canal, debemos usar un lector de *feeds* que permita dicha posibilidad.

Como ya comentamos anteriormente un *agregador* o lector de *feeds* podría ser además de un explorador web, un lector RSS instalado en el ordenador, un programa de correo electrónico o un lector online. Los lectores online son cómodos ya que no hay que instalarlos y su interfaz es fácil y familiar al ser de tipo web. Algunos de ellos son *Feedly*, *Inoreader* o *Newsblur*.

Una vez que accedemos con nuestra cuenta a uno de estos lectores, se nos presenta la posibilidad de suscribirnos a una serie de *feeds* preestablecidos, buscar *feeds* o indicar directamente el canal al que queremos suscribirnos con la opción “Añadir una suscripción”. Por ejemplo, podemos suscribirnos al canal de noticias de la Universidad de Granada (<https://www.ugr.es/en/node/4576>), momento a partir del cual el lector de *feeds* nos presentará el canal y sus noticias con una interfaz bastante agradable y con la posibilidad de usar muchas opciones que nos permiten gestionar los canales.

1.3. Un ejemplo.

Vamos a presentar un ejemplo que nos permita comparar *RSS* con *Atom*.

El sitio web <https://digibug.ugr.es/> contiene al final de su página tres pequeños iconos como los de la figura que sirven de enlace a la página de suscripción de las novedades. 

Esta página permite la suscripción al Repositorio Institucional de la Universidad de Granada tanto en formato *RSS* como *Atom*. Las páginas de notificación de novedades podemos verlas con un cliente o explorador web, aunque lo suyo es utilizar un lector apropiado. Cada cliente o explorador web mostrará estas páginas a su manera, ya que si la página no tiene una hoja de estilos propia, el cliente web usará la suya por defecto.

Si pinchamos en uno de estos enlaces obtendremos el código fuente de la página. Parte del código fuente de la página en formato *RSS* es el siguiente:

```
<?xml version="1.0" encoding="iso-8859-1"?>
<rss version="2.0">
  <channel>
    <title>Repositorio Institucional de la Universidad de Granada</title>
    <link>>https://digibug.ugr.es:443</link>
    <description>El repositorio digital DIGIBUG captura, almacena, indexa, preserva y distribuye materiales de investigación en formato digital.</description>
    <pubDate xmlns="http://apache.org/cocoon/i18n/2.1">Fri, 07 May 2021 02:33:09 GMT</pubDate>
    <item>
      <title>Calibrated prevalence of disabling chronic pain according to different approaches: a face-to-face cross-sectional population-based study in Southern Spain</title>
      <link>http://hdl.handle.net/10481/68366</link>
      <description>Calibrated prevalence of disabling chronic pain...</description>
      <guid isPermaLink="false">http://hdl.handle.net/10481/68366</ guid>
    </item>
    ...
  </channel>
</rss>
```

De igual forma, parte del código fuente de la página en formato *Atom* es este:

```
<?xml xmlns="1.0" encoding="iso-8859-1"?>
<feed xmlns="http://www.w3.org/2005/Atom" xmlns:dc="http://purl.org/de/elements/1.1.">
  <title>Repositorio Institucional de la Universidad de Granada</title>
  <link href="https://digibug.ugr.es:443" rel="alternate"/>
  <subtitle>El repositorio digital DIGIBUG captura, almacena, indexa, preserva y distribuye materiales de investigación en formato digital. Institucional de la Universidad de Granada</ subtitle >
  <id xmlns="http:// apache.org/cocoon/i18n/2.1">https://digibug.ugr.es:443</id>
  <updated>2021-05-07T08:35:36Z</updated>
  <dc:date>2021-05-07T08:35:36Z</ dc:date>
  <entry>
    <title>Perceptions of Teachers in Training on Water Issues and Their Relationship to the SDGs</title>
    <link href="http://hdl.handle.net/10481/68380" rel="alternate"/>
    <author>
      <name>Ruiz Garzón, Francisca</name>
    </author>
    <author>
      <name>Olmos Gómez, María Del Carmen</name>
    </author>
    <author>
      <name>Estrada Vidal, Ligia Isabel</name>
    </author>
    <id>http://hdl.handle.net/10481/68380</id>
    <updated>2021-05-07T08:12:09Z</updated>
    <content><![CDATA[Con algo de retraso (el verano es lo que tiene) publicamos el vídeo y el material de la décima charla gratuita de Autentia sobre]]>
    </content>
    <summary type="text">Perceptions of Teachers in Training on...</summary>
  </entry>
  ...
</feed>
```

2. Formatos.

En los códigos de los documentos *RSS* y *Atom* anteriores, por tratarse de archivos *XML*, podemos apreciar que comienzan con la misma cabecera en la que se indican el lenguaje *XML*, su versión y el código de caracteres que se está utilizando.

2.1. Formato RSS.

El formato *RSS*, permite describir las noticias de un canal (*channel*), del que se especifica un título (*title*), un enlace (*link*), que indica la página web donde se encontraría el canal competo, y una descripción (*description*) del canal. Opcionalmente, un canal puede tener otras etiquetas, tales como, *language* (para indicar el lenguaje usado en el texto), *webMaster* (para indicar un e-mail), *image* (para añadir una imagen), *pubDate* (para indicar la fecha de publicación del canal), etc.

A su vez, un canal está compuesto de un grupo de artículos (*items*) formados, como mínimo, por un título (*title*), un enlace a la URL donde se puede encontrar (*link*), y una descripción, es decir, un contenido del artículo (*description*). Además, puede tener otras informaciones, como fecha, autor, sitio de publicación, etc.

2.2. Formato Atom.

Al igual que en el formato *RSS*, los elementos de *Atom* son auto-explicativos pero la mayor diferencia que se puede apreciar entre ambos está en el nombre de algunas de las partes de los documentos. Por ejemplo, la etiqueta *feed* de *Atom* se corresponde con la etiqueta *channel* de *RSS*, y la etiqueta *Atom entry*, con *item* de *RSS*.

Las normas de *Atom* están definidas en la página web www.ietf.org/html.charters/atompub-charter.html.

El elemento raíz se denomina *feed*. y, a partir de él, se crea un encabezamiento con los detalles generales del canal de noticias para, posteriormente, colocar cada una de ellas secuencialmente.

El encabezamiento del *feed* dispone de un título (*title*), con un enlace (*link*) que indica la página web donde se encontraría el canal completo, un identificador único de recursos (*id*), una descripción (*subtile*) del canal, así como, otros elementos para indicar el autor, la fecha, logotipo, etc.

Cada artículo o entrada (*entry*), queda descrita mediante un identificador (*id*), un título (*title*) y un resumen de la noticia (*content*). Además se puede añadir la fecha de la noticia, el autor, etc.

2.3. Validación.

Normalmente no es necesario dominar los formatos anteriores. Existen herramientas que nos ayudan a editar un documento *RSS* o *Atom*. No obstante si queremos construirlo manualmente, lo más cómodo es usar un documento ya existente como una plantilla. Con objeto de comprobar si uno de estos documentos está bien formado podemos usar un explorador web, mientras que, para su validación existen herramientas específicas como la que tiene disponible el W3C, tanto para *RSS* como para *Atom*, en <http://validator.w3.org/feed/>. En <http://feedvalidator.org/> se pueden validar *feeds* pero sólo permite validar indicando la dirección web que contiene el documento a validar.

2.4. Editar feeds.

A la hora de crear un documento con formato *RSS* o *Atom* lo más cómodo es usar un editor específico para dicha tarea. Estos editores no solo permiten crear canales con sus artículos, sino que además permiten la gestión completa de varios canales simultáneamente y su publicación en el servidor web que queremos.

La empresa *mirabyte* ha desarrollado la aplicación *Feed Writer 2*. Existe una versión *free* que se puede usar bajo registro. Dicha versión debe ser usada para uso personal y no para un aprovechamiento empresarial o comercial. Se puede descargar de <http://download.mirabyte.com/feedwriter2.exe>.

Una vez instalada, al ejecutarla la aplicación lo primero que haremos será construir un nuevo archivo mediante la opción *New feed*. A continuación, elegiremos el formato *RSS* o *Atom* en sus diferentes versiones (elegiremos *RSS 2.0*) y un nombre para el archivo, por ejemplo *LaGaceta.xml*.

En el formulario que aparece seguidamente introduciremos los datos del canal imprescindibles: *title*, *description* y *link*. El resto de los datos, *language*, *date*, *logo*, etc., son opcionales.

Completada la cabecera, pulsaremos el botón *Feed Items* y se presentará un nuevo formulario para añadir las noticias al canal. El menú contextual del formulario permite gestionar *feeds*, es decir, añadir, copiar, borrar, mover, etc. Elegiremos *Add Item* para añadir un *feed* e introduciremos los datos obligatorios de este, *title*, *link* y *description*. En las otras pestañas del formulario podemos añadir más datos opcionales para el *feed*, tales como, *date*, *autor*, *guid*, etc.

El botón *Source Code* presenta el contenido del archivo en formato texto. En nuestro ejemplo, será el siguiente:

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

```

<rss version="2.0">
  <channel>
    <title>El saltabalates. La gaceta de Pelacañas.</title>
    <link>http://www.laperalimonera.es</link>
    <description>Acontecimientos de última hora</description>
    <language>es-es</language>
    <managingEditor>yo@pepe.es</managingEditor>
    <webMaster>Yoyo@pepe.es</webMaster>
    <lastBuildDate>Thu, 12 Aug 2010 18:17:29 +0200</lastBuildDate>

    <item>
      <title>Dinamitó la puerta de la casa</title>
      <link>http://www.laperalimonera.es/noticias/22.html</link>
      <description>Le dijeron que su puerta de seguridad era a prueba de bombas. Así que ni corto ni perezoso...</description>
      <pubDate>Thu, 12 Aug 2010 20:24:00 +0200</pubDate>
    </item>

    <item>
      <title>Se comió 67 higos chumbos.</title>
      <link>http://www.laperalimonera.es/noticias/21.htm</link>
      <description>Se levantó, se afeitó y, a continuación, y sin decir esta boca es mía se colocó por delante un saco de higos chumbos. Fue lo nunca visto.
    </description>
      <pubDate>Thu, 12 Aug 2010 19:48:30 +0200</pubDate>
    </item>

  </channel>
</rss>

```

El archivo *LaGaceta.xml* puede ser validado directamente desde la aplicación, mediante el botón *Validated Feed* ya que esta accede al validador del W3C en <http://validator.w3.org/feed/>.

Con un explorador web, o bien desde la misma aplicación con el botón *Preview Feed*, podemos presentar el archivo en pantalla. Dado que su aspecto no es muy adecuado, es aconsejable aplicarle una hoja de estilos para mejorar su diseño, o transformarlo a otro formato visual.

Como hemos estudiado anteriormente, el lenguaje *XSL* permite transformar un documento *XML* en otros tipos de documentos, entre ellos, en formato *HTML*, por lo que vamos a utilizarlo con objeto de obtener una web para nuestro canal de noticias. El archivo que definiremos para la transformación es *LaGaceta.xslt*:

```

<?xml version="1.0" encoding="UTF-8" ?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
  <xsl:preserve-space elements="*" />
  <xsl:template match="/">
    <html>
      <head>
        <title>
          <xsl:value-of select="rss/channel/title/text()" />
        </title>
      </head>
      <body>
        <h1>
          <xsl:value-of select="rss/channel/title/text()" />
        </h1>
        <h2>
          <xsl:value-of select="rss/channel/description/text()" />
        </h2>
        <a href="#">
          <xsl:value-of select="rss/channel/link/text()" />
        </a>
        <xsl:for-each select="rss/channel/item">
          <h3>
            <xsl:value-of select="title/text()" />
          </h3>
          <a href="#">
            <xsl:value-of select="link/text()" />
          </a>
          <p>

```

```
<xsl:value-of select="description/text()" />
</p>

<h4>
<xsl:value-of select="pubDate/text()" />
</h4>
</xsl:for-each>
</body>
</html>
</xsl:template>
</xsl:stylesheet>
```

Añadiremos al prólogo de nuestro archivo *LaGaceta.xml*, la etiqueta que lo enlaza a la hoja de estilos:

```
<?xml-stylesheet type="text/xsl" href="LaGaceta.xslt"?>
```

Si ahora abrimos el archivo *LaGaceta.xml* con un explorador web o bien usamos la opción *Preview Feed*, se presentará ya transformado en *HTML* con el siguiente aspecto:

El saltabalates. La gaceta de Pelacañas.

Acontecimientos de última hora

<http://www.laperalimonera.es>

Dinamitó la puerta de la casa

<http://www.laperalimonera.es/noticias/22.html>

Le dijeron que su puerta de seguridad era a prueba de bombas. Así que ni corto ni perezoso...

Thu, 12 Aug 2010 20:24:00 +0200

Se comió 67 higos chumbos.

<http://www.laperalimonera.es/noticias/21.htm>

Se levantó, se afeitó, y a continuación, y sin decir esta boca es mía se colocó por delante un saco de higos chumbos. Fue lo nunca visto.

Thu, 12 Aug 2010 19:48:30 +0200

El resultado de la transformación ha sido este archivo en formato *HTML*:

```
<html>
<head>
  <META http-equiv="Content-Type" content="text/html; charset=utf-8">
  <title>El saltabalates. La gaceta de Pelacañas.</title>
</head>
<body>
  <h1>El saltabalates. La gaceta de Pelacañas.</h1>
  <h2>Acontecimientos de última hora</h2>
  <a href="#">http://www.laperalimonera.es</a>
  <h3>Dinamitó la puerta de la casa</h3>
  <a href="#">http://www.laperalimonera.es/noticias/22.html</a>
  <p>Le dijeron que su puerta de seguridad era a prueba de bombas. Así que ni corto ni perezoso...</p>
  <h4>Thu, 12 Aug 2010 20:24:00 +0200</h4>
  <h3>Se comió 67 higochumbos.</h3>
  <a href="#">http://www.laperalimonera.es/noticias/21.htm</a>
  <p>Se levantó, se afeitó y, a continuación, y sin decir esta boca es mía se colocó por delante un saco de higochumbos. Fue lo nunca visto.</p>
```



```

<h4>Thu, 12 Aug 2010 19:48:30 +0200</h4>
</body>
</html>

```

Con objeto de asociarlo a una hoja de estilos *CSS* que dote a la web de propiedades que mejoren su diseño, añadiremos al archivo *LaGaceta.xlst* la etiqueta *link* con el archivo *CSS* que deberá usarse.

```

<?xml version="1.0" encoding="UTF-8" ?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
  <xsl:preserve-space elements="*" />
  <xsl:template match="/">
    <html>
      <head>
        <title>
          <xsl:value-of select="rss/channel/title/text()" />
        </title>
        <link rel="stylesheet" type="text/css" href="LaGaceta.css" media="screen" />
      </head>
      <body>
        . . . .

```

De esta forma, el archivo *HTML* resultante de la nueva transformación sería:

```

<html>
  <head>
    <META http-equiv="Content-Type" content="text/html; charset=utf-8">
    <title>El saltabalates. La gaceta de Pelacañas.</title>
    <link rel="stylesheet" type="text/css" href="LaGaceta.css" media="screen" />
  </head>
  <body>
    <h1>El saltabalates. La gaceta de Pelacañas.</h1>
    <h2>Acontecimientos de última hora</h2>
    . . . .

```

Como indica el atributo **href**, el archivo *CSS* que hemos usado, *LaGaceta.css*, es el siguiente:

```

body
{
    margin-left:1em;
}
h1
{
    font-size: 2em;
    font-family:Arial;
    color:Maroon;
    margin-bottom:-0.5em;
}
h3
{
    font-size: 1.2em;
    margin-top:1.5em;
    margin-bottom:-0.1em;
}
h2
{
    font-size: 1.5em;
    margin-bottom:0em;
    color:Gray;
}
h4
{
    font-size: 0.7em;
    margin-top:-1em;
}
p
{
    font-family:Verdana;
    font-size:0.9em;
    margin-top:0.5em;
}

```

Como resultado un navegador web mostrará el archivo *HTML* resultante de la transformación de esta forma:

El saltabalates. La gaceta de Pelacañas.

Acontecimientos de última hora

<http://www.laperalmonera.es>

Dinamitó la puerta de la casa

<http://www.laperalmonera.es/noticias/22.html>

Le dijeron que su puerta de seguridad era a prueba de bombas. Así que ni corto ni perezoso...

Thu, 12 Aug 2010 20:24:00 +0200

Se comió 67 higochumbos.

<http://www.laperalmonera.es/noticias/21.htm>

Se levantó, se afeitó, y a continuación, y sin decir esta boca es mía se colocó por delante un saco de higochumbos. Fue lo nunca visto.

Thu, 12 Aug 2010 19:48:30 +0200

Con este proceso nos hemos conseguido presentar el archivo *LaGaceta.xml* que contiene una fuente web para redifusión de noticias en formato RSS, en un formato adecuado gracias al uso de una transformación XSL a formato HTML, al cual le hemos asociado una hoja de estilos CSS.