

DIABD. Módulo de sindicación de contenidos

Práctica 1: Automatizar la creación de canales de sindicación

Profesor Manuel Blázquez Ochando

Introducción

Los canales de sindicación son archivos XML cuya actualización y regeneración se efectúa de manera automática. El propósito de esta práctica es conocer el método de regeneración y actualización de los contenidos en un canal de sindicación. Dicho de otra forma, cómo opera un sistema de publicación web, ya fuere un blog, wiki o CMS para mantener actualizado un canal de sindicación.

En la presente práctica se propone el desarrollo de un generador de canales de sindicación que integre todos los contenidos del CMS Joomla, así como cualquier otro contenido remitido fuera de Joomla.

Explicación:

1. Identificar la fuente de datos

Para generar un canal de sindicación con todos los contenidos de Joomla, se necesita conocer la estructura de tablas de la base de datos que lo soporta. Concretamente la tabla de contenidos principal se identifica como "**jos_content**". Para tener un conocimiento más exacto de los datos que se utilizarán es condición indispensable analizar todos los campos disponibles. Para la siguiente práctica se han seleccionado los siguientes:

Tabla principal: jos_content	
Nombre del campo	Descripción y características
id	Número correlativo identificador del contenido.
title	Título del contenido.
alias	URI del contenido a partir del título.
introtext	Resumen o texto introductorio.
fulltext	Texto o contenido completo.
catid	Categoría temática a la que corresponde el contenido. El tipo de campo es numérico, luego está relacionado con la tabla jos_categories

DIABD. Módulo de sindicación de contenidos

Práctica 1: Automatizar la creación de canales de sindicación

Profesor Manuel Blázquez Ochando

created	Fecha de creación del contenido.
created_by	Autor del contenido. El tipo de campo es numérico, luego está relacionado con la tabla jos_users
publish_up	Fecha de actualización del contenido.

Del análisis de los campos, se desprende que no todos los contenidos son de tipo de textual, ya que existen campos numéricos que indican que hay un dato relacionado en una tercera tabla. Este es el caso de los campos de categoría y de autor que dependen de las tablas **jos_categories** y **jos_users**.

Teniendo en cuenta que han aparecido tablas relacionadas, se necesita revisar la estructura de las mismas:

Tabla de categorías: jos_categories	
Nombre del campo	Descripción y características
id	Campo autonumérico, clave de la tabla categorías.
title	Denominación de la categoría.

Tabla de usuarios: jos_users	
Nombre del campo	Descripción y características
id	Campo autonumérico, clave de la tabla usuarios.
name	Nombre del usuario, en este caso, autor del contenido.

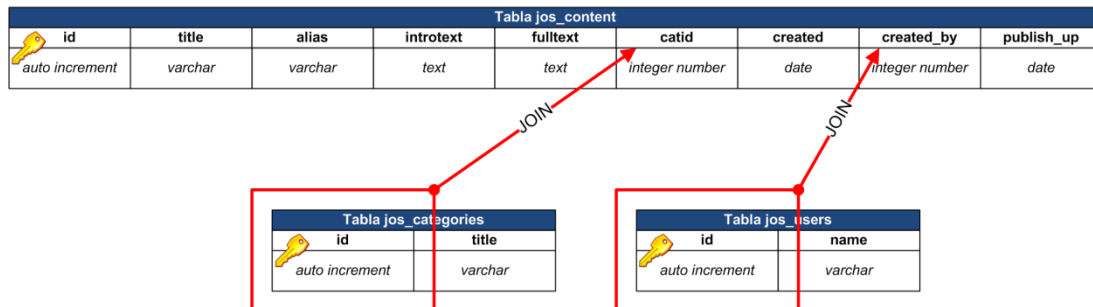
Para comprender mejor cuál es el esquema de datos que se necesita para llevar a efecto la práctica, hay que elaborar un cuadro de relaciones entre las tablas a fin de saber exactamente qué campos relacionados se requieren y poner en claro cuáles son los campos clave que permiten efectuar la relación en cada caso.

DIABD. Módulo de sindicación de contenidos

Práctica 1: Automatizar la creación de canales de sindicación

Profesor Manuel Blázquez Ochando

Relaciones de la tabla jos_content



Teniendo en cuenta estas circunstancias, se está en disposición de diseñar el programa generador y más concretamente la consulta SQL que permitirá extraer toda la información de los contenidos del CMS Joomla.

2. Consulta y extracción de datos

La fase de consulta y extracción de datos es la más importante, dado que posibilita el manejo y gestión de la información para ser plasmada posteriormente en un canal de sindicación. El objetivo por tanto, es consultar mediante SQL las tablas y campos propuestos. De hecho la primera y principal consulta que hay que efectuar es la siguiente:

Primera aproximación a la consulta SQL

```
SELECT * FROM jos_content
```

Esta primera aproximación permite seleccionar todos los campos de la tabla **jos_content** para recuperarlos posteriormente. Pero como anteriormente se ha mencionado hay datos como la categoría y el nombre del autor del contenido que están relacionados mediante su identificador, puesto que es el único dato que se obtendría.

Por tanto para llevar a cabo una consulta relacionada con las tablas **jos_categories** y **jos_users**, se necesita emplear la instrucción JOIN ON:

DIABD. Módulo de sindicación de contenidos

Práctica 1: Automatizar la creación de canales de sindicación

Profesor Manuel Blázquez Ochando

Sintaxis de JOIN ON

SELECT campos de la tabla principal y relacionadas **FROM** tabla principal **JOIN** tabla relacionada **ON** campo relacionado de la tabla principal = campo clave

JOIN permite unir una tercera tabla a la selección que originalmente se hace con **FROM**. Dicha unión requiere de un campo clave por el cual filtrar el número id que las relaciona. Ello se expresa mediante la cláusula **ON** de forma que define el campo clave por el cual se relacionan las tablas. Éste método puede repetirse tantas veces sea necesario, según tantos campos clave se dispongan en la tabla de contenidos.

Segunda aproximación a la consulta SQL

```
SELECT jos_categories.title, jos_users.name, jos_content.title FROM jos_content JOIN
jos_categories ON jos_content.catid = jos_categories.id JOIN jos_users ON
jos_content.created_by = jos_users.id
```

En esta segunda aproximación se extraen el nombre de la categoría (`jos_categories.title`), el nombre del autor (`jos_users.name`) y el título del contenido (`jos_content.title`), marcados en amarillo. Por otro lado se selecciona la tabla principal de contenidos **jos_content**, marcada en rojo. A continuación se utiliza **JOIN** para establecer la tabla relacionada de categorías **jos_categories**, marcada en verde y la cláusula **ON** para establecer el campo clave entre **jos_content** y **jos_categories**, que en este caso es `jos_content_catid = jos_categories.id`, marcado en color magenta.

Dado que hay una segunda tabla relacionada, al finalizar la primera instrucción **JOIN ON**, se vuelve a reproducir de nuevo con la tabla de usuarios, en la que se encuentra el nombre o nombres de los autores de los contenidos. Por ello, marcado en verde se declara el nombre de la tabla de usuarios (`jos_users`) y en magenta se establece el campo clave de la tabla principal (`jos_content.created_by`), que permite establecer la relación oportuna con el campo clave de usuarios (`jos_users.id`).

Pero con esta última aproximación no se obtienen todos los campos que se necesitan, por lo que se necesita ampliarlos desde **SELECT**, tal y como muestra la siguiente tabla.

Tercera aproximación a la consulta SQL

```
SELECT jos_categories.title, jos_users.name, jos_content.title, jos_content.alias,
jos_content.introtext, jos_content.fulltext, jos_content.created, jos_content.publish_up
FROM jos_content JOIN jos_categories ON jos_content.catid = jos_categories.id JOIN
jos_users ON jos_content.created_by = jos_users.id
```

DIABD. Módulo de sindicación de contenidos

Práctica 1: Automatizar la creación de canales de sindicación

Profesor Manuel Blázquez Ochando

Con la tercera aproximación se llega a seleccionar todos los campos necesarios para efectuar la consulta de las tres tablas, pero no todo está en orden.

¿Qué ocurre si existen dos campos cuya denominación es la misma? Es el caso de los campos **title**. Ya que no es lo mismo el **title** de la categoría que el **title** del propio contenido, se necesita una solución que permita identificar de forma unívoca uno y otro. Para ello se emplea la cláusula **AS** que sirve para crear un campo alias o avatar que diferencia al menos uno de los dos.

Consulta SQL definitiva

```
SELECT jos_categories.title AS category, jos_users.name, jos_content.title, jos_content.alias,
jos_content.introtext, jos_content.fulltext, jos_content.created, jos_content.publish_up
FROM jos_content JOIN jos_categories ON jos_content.catid = jos_categories.id JOIN
jos_users ON jos_content.created_by = jos_users.id ORDER BY jos_content.publish_up DESC
```

Tal y como se muestra en la tabla anterior, la consulta SQL definitiva, contempla la creación de un campo avatar denominado **category**. De esta forma el título de la categoría no se confundirá con el título de la tabla de contenidos.

Como último detalle se marca en color magenta la instrucción para ordenar los contenidos según su fecha de actualización, de forma tal que cuando se introduzcan nuevos contenidos, aparezcan referidos en primer lugar.

3. Desarrollando el programa generador

Dispuesto el código SQL para la consulta de las tablas, el siguiente paso es la confección del programa generador propiamente dicho. Antes de comenzar este trabajo es necesario determinar cuál será el formato de sindicación en el que se van a plasmar todos los contenidos. (Para esta práctica se ha elegido el formato **Atom**).

Determinados estos aspectos, se seguirán los siguientes pasos:

1. Crear una carpeta en la cuenta del servidor. Dicha carpeta se denominará “**sindicacion**”.
2. Aplicar permisos de lectura, escritura y ejecución CHMOD 777 utilizando para ello el programa WINSXP.
3. Crear un archivo **config.php** que contendrá los datos de conexión al CMS Joomla.
4. Crear un archivo **generador-atom.php** que contendrá el programa generador.

DIABD. Módulo de sindicación de contenidos

Práctica 1: Automatizar la creación de canales de sindicación

Profesor Manuel Blázquez Ochando

3.1. El archivo config.php

El archivo config.php permite definir los datos de nombre de usuario y contraseña de MySQL así como nombre del servidor. También se añade el nombre de la base de datos de Joomla en la variable `$database` que será utilizada posteriormente en el programa generador.

Archivo: config.php
<pre><?php \$con = mysql_connect('localhost', 'root', 'root') or die ('error: no se pudo conectar a mysql'); \$database = 'joomla'; ?></pre>

3.2. El archivo generador-atom.php

Como se ha explicado en la tarea1 así como en los materiales de clase, los canales de sindicación se componen de dos apartados bien diferenciados, por un lado el canal propiamente dicho y por otro lado los ítems que contienen los contenidos o artículos del mismo

Un programa generador debe ser capaz de crear ambos apartados e incluirlos en un único archivo XML que contiene el canal de sindicación.

Teniendo en cuenta que se dispone de la estructura del formato de sindicación Atom, el problema más difícil de resolver es la creación de un archivo XML desde el propio programa.

En este sentido existen funciones para la apertura, edición y cierre de archivos en PHP. Concretamente las funciones **fopen**, **fwrite** y **fclose**.

Funciones de edición de archivos en PHP	
Nombre de la función	Descripción
<code>\$file = fopen("nombre-archivo.xml", "modo de escritura");</code>	La función fopen, permite abrir o crear si no existe un archivo. También puede definir el modo de escritura del archivo. El modo de escritura W resulta común para escribir y regenerar en todo caso el archivo XML.
<code>fwrite(\$file, "contenido");</code>	La función fwrite, define en un primer argumento en amarillo, el archivo en el que se escribe y en un segundo argumento en verde el contenido que se pretende escribir.

DIABD. Módulo de sindicación de contenidos

Práctica 1: Automatizar la creación de canales de sindicación

Profesor Manuel Blázquez Ochando

<code>fclose(\$file)</code>	Se emplea para cerrar el archivo
-----------------------------	----------------------------------

Aclarados estos aspectos el primer paso del desarrollo del programa es crear el archivo XML que constituye el canal de sindicación, así como escribir la estructura XML correspondiente a los elementos que lo describen:

Primer avance del generador
<pre><?php \$file = fopen("canal-atom-practica2.xml","w"); fwrite(\$file,"<?xml version='1.0' encoding='UTF-8'?> <feed xmlns='http://www.w3.org/2005/Atom'> <id></id> <title>escribe aqui el título de tu canal</title> <updated></updated> <subtitle>describe brevemente tu canal</subtitle> <author><name>escribe tu nombre</name></author> ");</pre>

Nótese que en este primer avance del generador, existen etiquetas vacías como el `<id></id>` y `<updated></updated>`. Esto se debe a que el identificador es la URI del propio canal de sindicación, dato que aunque puede ser introducido manualmente, conviene también automatizar. Por otro lado la fecha de actualización debe mantener un formato normalizado ISO 8601, aspecto para el cual también existe una solución automática:

Soluciones para la automatización del URI del canal y la fecha completa	
Solución	Descripción
<code>\$datetime = date(c);</code>	La función date de php, permite crear automáticamente una fecha estándar según el reloj del servidor. Introduciendo el valor "c" se selecciona el formato ISO8601.
<code>\$uri="http://\$_SERVER[HTTP_HOST]/sindicacion/canal-atom-practica2.xml"</code>	La forma de obtener la URI del canal de sindicación puede elaborarse a mano o semi-automatizada, mediante la construcción del valor en una variable. Hay que destacar el papel de la variable constante <code>\$_SERVER[HTTP_HOST]</code> que permite obtener el nombre del servidor para componer la URL final.

Efectuando estos cambios se obtiene el siguiente resultado:

DIABD. Módulo de sindicación de contenidos

Práctica 1: Automatizar la creación de canales de sindicación

Profesor Manuel Blázquez Ochando

Segundo avance del generador

```
<?php
$datetime = date(c);
$uri="http://$_SERVER[HTTP_HOST]/sindicacion/canal-atom-practica2.xml"

$file = fopen("canal-atom-practica2.xml","w");
fwrite($file,"<?xml version='1.0' encoding='UTF-8'?>
<feed xmlns='http://www.w3.org/2005/Atom'>
    <id>$uri</id>
    <title>escribe aqui el título de tu canal</title>
    <updated>$datetime</updated>
    <subtitle>describe brevemente tu canal</subtitle>
    <author><name>escribe tu nombre</name></author>
");
```

De esta forma los datos esenciales del canal Atom pueden ser escritos correctamente, no obstante la edición del programa no ha terminado, ya que aún queda desarrollar la parte del programa que genera los contenidos. Para ello se ejecutará la consulta SQL anteriormente desarrollada (marcada en amarillo en la siguiente tabla) para obtener un **array asociativo** (marcado en verde) que permita imprimir los resultados, o mejor dicho, escribirlos en el archivo. Para que este proceso funcione de forma iterativa con cada uno de los resultados registrados en el array, es preciso contar con un bucle del tipo **while**, que ejecutará la instrucción de escritura tantas veces como filas de resultados se hayan obtenido.

Al finalizar la ejecución del bucle **while** es necesario escribir la etiqueta de cierre del canal Atom `</feed>`, para ello se utiliza a función `fwrite`, y acto seguido `fclose`, para cerrar la edición del archivo.

El resultado de programar estos aspectos es el que se muestra en la siguiente tabla.

Archivo: generador-atom.php

```
<?php

$datetime = date(c);
$uri="http://$_SERVER[HTTP_HOST]/sindicacion/canal-atom-practica2.xml"

include("config.php");

$file = fopen("canal-atom-practica2.xml","w");
fwrite($file,"<?xml version='1.0' encoding='UTF-8'?>
<feed xmlns='http://www.w3.org/2005/Atom'>
    <id>$uri</id>
    <title>escribe aqui el título de tu canal</title>
    <updated>$datetime</updated>
    <subtitle>describe brevemente tu canal</subtitle>
    <author><name>escribe tu nombre</name></author>
");
```


DIABD. Módulo de sindicación de contenidos

Práctica 1: Automatizar la creación de canales de sindicación

Profesor Manuel Blázquez Ochando

```
mysql_select_db("$database", $con);
$sql = "SELECT jos_categories.title AS category, jos_users.name, jos_content.title,
jos_content.alias, jos_content.introtext, jos_content.fulltext, jos_content.created,
jos_content.publish_up FROM jos_content JOIN jos_categories ON jos_content.catid =
jos_categories.id JOIN jos_users ON jos_content.created_by = jos_users.id ORDER BY
jos_content.publish_up";
$resultado = mysql_query($sql,$con);
while ($fila = mysql_fetch_assoc($resultado)) {

    fwrite($file, "
    <entry>
        <id>http://$_SERVER[HTTP_HOST]/joomla/$fila[alias]</id>
        <title>$fila[title]</title>
        <updated>$fila[publish_up]</updated>
        <author><name>$fila[name]</name></author>
        <content type='xhtml'>$fila[introtext] - $fila[fulltext]</content>
        <category term='$fila[category]'/>
        <published>$fila[created]</published>
    </entry>
    ");
}

fwrite($file, "</feed>");
fclose($file);

?>
```

Una vez generados los archivos y dispuestos en la carpeta sindicación del servidor, sólo es necesario introducir la URL:

<http://thor.eubd.ucm.es/~diabycXX/sindicacion/generador-atom.php>

Al cargar el programa se ejecutará todo el proceso de generación, dando como resultado un archivo denominado **canal-atom-practica2.xml**, que constituye en sí mismo el canal de sindicación.

Práctica:

1. Probar todos los puntos de la explicación en el servidor, dado que constituyen la explicación de la práctica y verificar que se ha creado el archivo XML. Se comprobará que todos los archivos fueron creados y funcionan adecuadamente.
2. Enlazar el nuevo canal de sindicación a la página de inicio de Joomla. Escribe qué instrucción se debe introducir en la cabecera de la página para que sea reconocido el canal por el navegador web.

DIABD. Módulo de sindicación de contenidos

Práctica 1: Automatizar la creación de canales de sindicación

Profesor Manuel Blázquez Ochando

Respuesta de la pregunta 2:

3. Introduce un nuevo contenido en Joomla y vuelve a cargar el generador, para comprobar que el contenido ha sido actualizado también en el canal de sindicación. Copia y pega en la siguiente tabla el código fuente del último ítem.

Respuesta de la pregunta3:

4. Crea un segundo generador de canales de sindicación, adaptado al formato RSS2.0 e introduce su código fuente en la siguiente tabla.

Respuesta de la pregunta4: