



Facultad 10

- Temas:**
- 1- Que es un repositorio de documentos, uso frecuentes.
 - 2- Uso específico para un sistema de gestión documental
 - 3- Estándar para java jsr -170. Facilidades, usabilidad.

Autora : Ing Velmour Muñoz Casals

1- Qué es un repositorio de documentos?

Repositorio de documentos de archivo significa una oficina de documento, una biblioteca, un museo o cualquier otra institución que custodia documentos que están abiertos para la inspección pública. (Debe observarse que en el Estándar Británico 5454 'repositorio' significa únicamente el lugar de almacenamiento).

Un repositorio es un sitio centralizado donde se almacena y mantiene información digital, habitualmente bases de datos o archivos informáticos.

El origen de la palabra española repositorio deriva del latín **repositorium**, que significaba armario, alacena. Este significado se generalizó (por metonimia) en español y es recogido en el Diccionario de la Real Academia (DRAE) como: "**Lugar donde se guarda algo**", y de ahí se aplicó al léxico específico de la informática para designar los repositorio de información digital.

Los repositorios están preparados para distribuirse habitualmente sirviéndose de una red informática como Internet o en un medio físico como un disco compacto. Y pueden ser de acceso público, o pueden estar protegidos y necesitar de una autenticación previa. Los repositorios más conocidos son los de carácter académico y los institucionales.

A diferencia de los ordenadores personales o de las PC de escritorio, los repositorios suelen contar con sistemas de Backup y mantenimiento preventivo y correctivo, lo que hace que nuestra información se pueda recuperar en el caso que nuestra máquina o PC quede inutilizable.

Objetivo de crear y mantener un repositorio de documentos

Lograr un establecimiento permanente para el cuidado y la inspección pública de documentos, fundamentado en una gestión sólida y en firmes garantías financieras.

Una manera de comprender los repositorios, es imaginar una combinación entre una biblioteca digital y un buscador como Yahoo o Google, pero mucho más sofisticado que ambos.

Consideraciones en la implementación de un repositorio.

- Para obtener información relevante, los usuarios no requieren conocer previamente la ubicación de los datos.
- Los nombres de tablas no debiesen ser extraños ni llevar a la confusión.
- Las consultas no debiesen acceder una gran cantidad de tablas.
- Un repositorio requiere la implementación de una capa de semántica para crear vistas comerciales que sean comprensibles y que almacene las reglas de join y los controles de los recursos.

Beneficios de un repositorio bien implementado

“Un director de recursos de una compañía necesita los nombres y número de empleados de todos aquellos que han estado en la compañía más de 5 años y ganaron más de \$15.000.000 por año. La información reside en distintas tablas. Con un repositorio, el director sencillamente formularía una consulta usando la información de la vista de recursos humanos. Sin el repositorio, el director necesitaría saber la ubicación de la información, cómo descifrar el nombre de tablas y ficheros (EMPEN en vez de "Nombre Empleado" o EMPEN en vez de "Codigo Empleado,"por ejemplo), y cómo realizar los joins y demás operaciones SQL antes de que pueda obtener la información”.

- Fácil acceso a los datos corporativos.
- Menor soporte y formación.
- Mejora del rendimiento debido a la optimización de joins de tablas.
- Resultados precisos, debido a que los usuarios no tienen que determinar los joins.

Ejemplo de Repositorios y su uso

Los grandes repositorios son “distribuidos” y la tendencia actual está orientada a crear enormes redes de repositorios locales. ARIADNE, por ejemplo, funciona en base a “knowledge pool” locales y centrales, interconectados y sujetos a un mismo estándar. El proyecto “eduSource” creará un repositorio pan-canadiense, integrando todos los repositorios y colecciones (más pequeñas que los repositorios y especializadas) de ese país. El esfuerzo integrador y globalizador, que domina el ambiente, se aprecia claramente en este proyecto, al rastrear entre sus componentes. Entre los repositorios que se integrarán aparece Alexandria, que a su vez se presenta como “el primer repositorio nacional de objetos educativos” y que corresponde a un prototipo de investigación resultante de un proyecto desarrollado por CAREO (Campus Alberta Repository of Educational Objects).

2- Uso específico para un sistema de gestión documental

Gestión documental: Es una herramienta que está orientada a la catalogación y recuperación de contenidos grandes (documentos). Incorpora a sus prestaciones, las capacidades necesarias para gestionar los procesos que crean, almacenan, tratan y presentan información, en entorno digital.

Repositorio de documentos de archivo: Significa una oficina de documento, una biblioteca, un museo o cualquier otra institución que custodia documentos que están abiertos para la inspección pública.

La relación que existe entre la gestión documental y los repositorios de documentos es que el repositorio es una capacidad actual de los gestores de contenido que son todos los procedimientos y procesos involucrados en la agregación, transformación, *catalogación*, agrupación, autorización, presentación y distribución de información útil para nuestros propósitos.

Metadatos: (del griego *μετα*, *meta*, «después de» y latín *datum*, «lo que se da», «dato»), literalmente «sobre datos», son datos que describen otros datos. En general, un grupo de metadatos se refiere a un grupo de datos, llamado *recurso*.

El concepto de metadatos es análogo al uso de índices para localizar objetos en vez de datos. Por ejemplo, en una biblioteca se usan fichas que especifican autores, títulos, casas editoriales y lugares para buscar libros. Así, los metadatos ayudan a ubicar datos.

Para varios campos de la informática como la recuperación de información o la web semántica, los metadatos son un enfoque importante para construir un puente sobre el intervalo semántico.

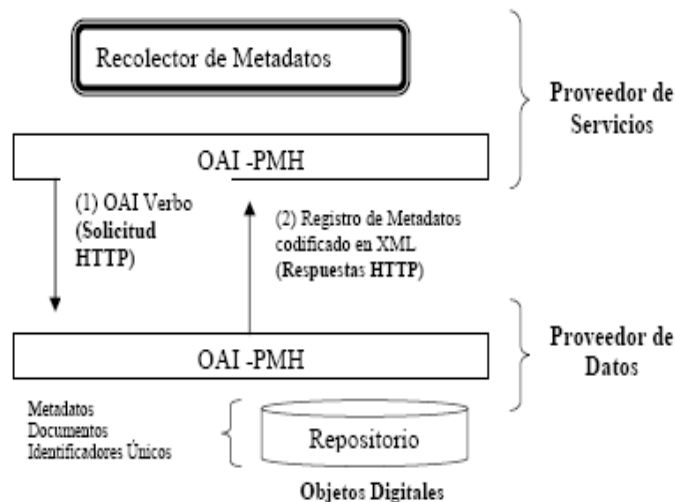


Figura 2. Modelo de transacciones OAI-PMH

El protocolo de recolección de metadatos (OAI-PMH), es el mecanismo de trabajo que debe ser implementado en los sistemas de información documental para poder comunicarse entre si (interoperabilidad), este protocolo define todos los “comandos ó verbos” necesarios para poder recolectar metadatos de forma automática y normalizada. Este protocolo define toda su funcionalidad en solo seis (6) verbos básicos, y algunos parámetros complementarios. En el caso de REDINED para utilizar este protocolo se debe invocar desde:

Verbo	Función
Metadatos acerca del repositorio	
Identify	Descripción del repositorio
ListMetadataFormats	Lista de todos los tipos de metadatos soportados en el repositorio
ListSets	Lista de las particiones lógicas de información en el repositorio
Verbos de recolección de metadatos	
ListIdentifiers	Lista de los identificadores OAI únicos existentes en el repositorio
ListRecords	Lista de n registros de metadatos
GetRecord	Lista de un solo registro de metadatos

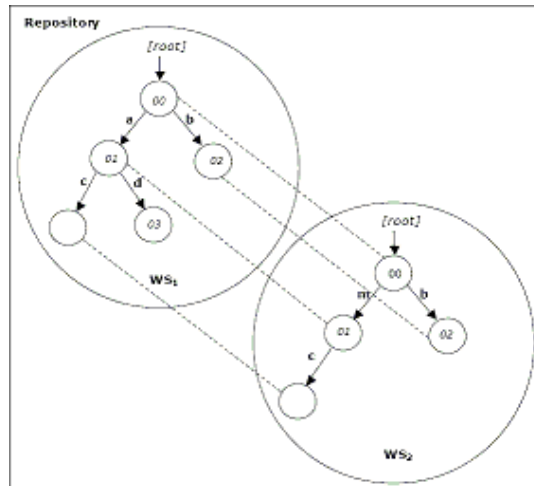
Tabla 3. Verbos del protocolo OAI-PMH

3- Estándar para java jsr – 170. Facilidades, usabilidad.

Esta especificación (JSR 170) es un paso hacia delante de los fabricantes de sistemas de gestión de contenidos (CMS): IBM, Vignette, BEA, Oracle, Day Software, Fujitsu, etc. La intervención de los más importantes ha dado esperanzas a que la especificación sea implementada por todos los fabricantes y desarrolladores Open Source. La implantación de esta especificación facilitará y abaratará el desarrollo de webs enterprise y de las herramientas necesarias para ello.

Arquitectura

La especificación dibuja un repositorio compuesto por árboles de nodos inter-relacionados (workspaces), donde cada nodo tiene una serie de propiedades y de nodos hijo, además de relaciones (opcionales) entre nodos.



Este diseño, en principio, parece lo suficientemente flexible como para encajar casi cualquier otro diseño jerárquico, que es el diseño normal en cualquier CMS. El problema se presenta cuando introducen el término Mixin Node Types (mixinTypes), que introduce un nivel más de complejidad al diseño, ya que indica las funcionalidades de cada uno de los nodos de forma particular. De serie vienen 3: referenciable, versionable y bloqueable. La problemática llegará cuando queramos exportar un repositorio de un proveedor de CMS a otro. Si en el CMS de destino no tenemos los mixinTypes del de origen, ¿qué pasará? ¿Podremos exportar también los mixinTypes? ¿Deberemos reprogramar las clases que representen al mixinType?

Escape de caracteres en el XML exportado

Debido al problema de que el nombre de un nodo no ha de ser un nombre XML, se ha tenido que introducir una forma de 'escapar' los nombres de nodos bastante extraña (6.4.3): "My Documents" -> "My_x0020_Documents"

En este aspecto, queda la especificación un poco rara, ya que necesitaremos una librería específica para tratar este tipo de codificaciones. Lo sorprendente es que Vignette, estando en el comité de expertos, no sugiriera su solución al respecto (se puede elegir el campo que será el mostrado para visualizar el nombre de un contenido).

Este tipo de codificación también se utiliza para escapar contenidos multivaluados (que se separan por espacios).

Localización (i18n)

Se echa de menos el mixinType 'localizable' para poder tener traducciones de manera inherente al diseño, sin tener que hacer luego un esfuerzo sobrehumano, utilizando las relaciones entre nodos para modelarlas.

Este es un problema común a todos los CMS que conozco (tampoco conozco muchos), y es incomprensible que CMS enterprise no tengan un buen soporte para Webs Enterprise, que se muestran en varios idiomas, tienen ficheros traducidos, etc.

Conclusión

Por esta serie de detalles, la JSR 170 puede que se utilice más como herramienta de exportación/importación básica que como plataforma sobre la que trabajar.

Referencias Bibliográficas

<http://es.wikipedia.org/wiki/Repositorio>

http://eprints.rclis.org/archive/00009841/01/Ponencia_COngreso.pdf

http://www.dominpe.com/articulos/articulos/jsr_170:_java_content_repository_jcp.html

<http://blogdrake.net/node/5701>

<http://www.alfresco.com/es/products/solutions/ecm/repository/>