

Automatizando la actualización de colecciones en GSDL

- Mariana Pichinini
- Enrique Merle

Memoria Académica

BIBHUMA Biblioteca Prof. Guillermo Obiols

Facultad de Humanidades y Ciencias de la Educación

Universidad Nacional de La Plata

Quiénes somos

- Memoria Académica, inicio 2006
- Aprox. 5000 recursos digitales y 10000 referencias
- Nos interesa controlar la calidad de la presentación final de los recursos (html)
- Se incluye una carátula a los pdfs de descarga para identificación y referenciación.
- Incluye tanto referencias como recursos a texto completo

A close-up photograph of a dandelion seed head in the process of blowing away, with many seeds floating in the air. The background is a soft, out-of-focus blue and green.

Memoria Académica

compartimos lo que sabemos

FaHCE-UNLP

Biblioteca digital existente

- GSDL v.2.82 producción v.2.84/.85rc desarrollo
- 9 colecciones
- Recursos textuales (documentos Word y html)
- Recursos descargables fuera de GSDL (PDF)
- 6 colecciones con metadatos gestionados fuera de Greenstone en bases CDS/ISIS
- 3 colecciones con metadatos gestionados por GSDL

Decisiones iniciales

- Gestión de recursos fuera del contexto de GSDL (sin uso de GLI)
 - Bases para gestión de metadatos, permite reutilización
 - Sólo colecciones acotadas con edición de plantillas en texto plano para creación de metadatos
 - Editor de texto configurable utf-8 (Notepad++)
 - Plantilla metadata.xml
- Import.pl y buildcol.pl por línea de comandos

Flujo de trabajo (procesos técnicos)

- 4 personas procesando recursos digitales
- Entre 4 y 5 recursos procesados diariamente
- Etapas del proceso:
 - Registro mínimo en base de datos correspondiente
 - Edición y conversión recurso entrante
 - Creación de carpeta definitiva del recurso
 - Registro completo en BD (metadatos y referencias)

Flujo de trabajo (administración)

- Ejecución de actualización de colecciones GSDL en servidor de trabajo
- Control de logs (registro de errores)
- Solicitud chequeo operadores en Sitio secundario GSDL (no público)
- Nueva ejecución según requerimientos de operadores ante correcciones
- Publicación en Sitio definitivo primario ante visto bueno

Problemas de la actualización manual

- Serie de procesos interdependientes que configuran una estructura compleja.
- La variedad de herramientas necesarias (basadas en Windows, Linux, y DOS de 16 bits), hace las actualizaciones muy sensibles a la correcta ejecución de cada paso.
- Todo esto termina consumiendo muchas horas de trabajo semanales para mantener las colecciones actualizadas.

Automatización del proceso

- Varios guiones (scripts) en bash
 - Parametrizaciones en archivos de texto
 - Llamado de formatos auxiliares externos
- En desarrollo front-end para descentralizar todo el procedimiento
 - Permisos de actualización discriminados permitirá que cada operador mantenga las colecciones a su cargo, desde el principio hasta la publicación final.
 - Se registran todas las operaciones lo cual permitirá rastrear problemas o errores.

Herramientas de consola

- Se adoptó un enfoque *bottom-up* (de abajo hacia arriba) para encarar la solución
- Cada solución puntual se articuló en una aplicación única que controla cada paso de ejecución a medida que va completando el procedimiento.
- El carácter secuencial del proceso de actualización facilita su automatización.
- Si un paso falla, el proceso completo se aborta sin consecuencias.
- En caso de necesidad puede ejecutarse individualmente cada paso.

Glosario BCERS (BCRS Collection Rebuilding System)

- **Bcrsys** – Aplicación principal que llama a **bcrtool** y **bcroc** en un bucle, respalda todo, y luego activa el proceso de importación de GSDL.
- **Bcrtool** – copia las colecciones a su destino de procesamiento recreando el árbol de carpetas, usando **bcrcopy**, controla y elimina duplicados de versiones con **bcrdup** y activa **bcrisis**.
 - **Bcrisis** – convierte bases ISIS en formato DOS 16 a ELF 64; vuelca el contenido a metadata.xml según carpetas y lo convierte a UTF-8 (original en iBM850).
- **Bcroc** – genera una versión para OAI de una colección dada.

Aplicación gráfica en desarrollo



BGDC™ Gestión de Colecciones

/servidor/procesamiento/4.LISTOS/B.Memoria

CONECTADO

Agente

Contraseña

CONECTAR

LISTAR

EDITAR

LISTAR METODOS

Enlaces relacionados

Espacio para publicidad

eventos PUBLICAR

tesis PUBLICAR

te.363 Aprobar Rechazar Sin verificación

te.363.html Aprobar Rechazar Deshacer

te.363.pdf Aprobar Rechazar Deshacer

te.382 Aprobar Rechazar Sin verificación

te.393 Aprobar Rechazar Sin verificación

te.400 Aprobar Rechazar Sin verificación

te.421 Aprobar Rechazar Sin verificación

progra PUBLICAR

arti PUBLICAR

planes PUBLICAR

convenio PUBLICAR

proyecto PUBLICAR

norma PUBLICAR

libros PUBLICAR

Finalmente

- Fue necesario conjugar diferentes tecnologías de distinto nivel de actualidad, lo que constituía uno de los núcleos del problema.
- El otro objetivo fue no duplicar trabajo y optimizar tiempos
- Individualizando y enfrentando cada problema por separado se obtuvieron soluciones sencillas y puntuales que luego se conectaron entre sí para lograr lo que buscábamos:
- Fama y fortuna..... y menos trabajo tracción a sangre.

Gracias por su atención

<http://www.memoria.fahce.unlp.edu.ar>

mariana@fahce.unlp.edu.ar

everness@fahce.unlp.edu.ar