

XII ASAMBLEA DE REBIUN

San Sebastián / Donostia, 3-5 de noviembre de 2004

PROPUESTA PARA LA INSERCIÓN DE HIPERENLACES DESDE EL CATÁLOGO COLECTIVO DE REBIUN A LOS REGISTROS CORRESPONDIENTES EN LOS CATALOGOS DE LAS BIBLIOTECAS PARTICIPANTES

GRUPO DE ACCESO AL DOCUMENTO

CSIC. Coordinación de Bibliotecas
Biblioteca de la Universitat de Barcelona
Biblioteca de la Universidad Carlos III
Biblioteca de la Universidad de Extremadura
Biblioteca de la Universidad de Jaén
Biblioteca de la Universitat Jaume I
Biblioteca de la Universidad de La Rioja
Biblioteca de la Universidad de Lleida
Biblioteca de la Universidad Politécnica de Madrid
Biblioteca de la Universitat Rovira i Virgili
Biblioteca de la Universidad de Santiago de Compostela
Biblioteca de la Universidad de Valladolid
Biblioteca de la Universidad de Zaragoza

Coordinación: Universidad de Pública de Navarra

SUPERVISADA POR EL GRUPO DE CATÁLOGO COLECTIVO

OBJETIVO:

Visualizar desde el catálogo colectivo de REBIUN la información sobre disponibilidad de las publicaciones resultantes de una búsqueda.

Precedentes

Uno de los engorros con los que se encuentra cualquier "localizador" de publicaciones de un servicio de Acceso al documento al consultar el catálogo colectivo de REBIUN es que éste no facilita acceso a las colecciones locales (salvo en el caso de las bibliotecas que utilizan como sistema el suministrado por la empresa que tiene subcontratado el servicio de Catálogo colectivo para REBIUN). Por este motivo, a los "localizadores" no les queda más remedio que, una vez encontrada una publicación en el catálogo colectivo, repetir la consulta en el (o los) catálogo(s) de la(s) institución(es) de que se trate, bajo su propia interfaz, requiriendo retectar la consulta y, en ocasiones, incluso pasar por varias pantallas de acceso al catálogo. De hecho, cuando este trabajo no se realiza y se envía una petición sin ese "chequeo" se corre el riesgo de que la respuesta sea finalmente negativa, porque hay algún error, la obra no es prestable, no se dispone de ese año o número de esa revista, o la obra se encuentra prestada en esos momentos. Nótese que más del 20% de las peticiones que nos hacemos entre Bibliotecas REBIUN reciben respuesta negativa, que casi siempre podría haberse evitado con este chequeo.

Esta situación provoca, o el incremento de las respuestas negativas referido anteriormente si no se realiza el chequeo, o una enorme desazón y el consumo, en una tarea absolutamente nada atractiva, de un tiempo innecesario (piénsese que por cada 100.000 consultas al catálogo para PI el conjunto de REBIUN pierde al menos 200.000 minutos, más de 3.300 horas al año, más de dos jornadas completas anuales de trabajo al año).

Desde el Grupo de PI se ha venido pidiendo desde hace tres años que este asunto se solucionase. Con el fin de contribuir a lograrlo, se ha pasado a preparar una propuesta concreta, de manera que en vez de pedir que se solucione podamos facilitar, si no la vía real para hacerlo, al menos una aproximación eficiente que permita abordar el problema de una vez y superarlo, que sea compartida, además, con el grupo de Catálogo colectivo.

Resultado que se desea conseguir

Que, situados en un registro bibliográfico del CC de REBIUN, pinchando en el acrónimo correspondiente a una biblioteca poseedora de esa publicación, se abra una nueva ventana del navegador en la que se vea el registro correspondiente a esa publicación en el catálogo local de la institución en cuestión.

Problema a examinar más adelante

En la mente de todos ha venido estando un problema que puede haber contribuido a no haberse abordado el asunto.

¿Deben las URL ser suministradas desde origen en los propios registros MARC y conservarse en el proceso de deduplicación para luego añadirse a los registros concentrados del catálogo?

¿Deben construirse por la empresa proveedora del CC en el proceso de aceptación de registros?

¿Deben construirse sólo en el momento de ser "solicitado" al pinchar un acrónimo?

Pensamos que este es un problema posterior y que antes deben zanjarse otros aspectos imprescindibles a cuya luz posiblemente éste quede contestado.

Pasos imprescindibles o hitos

1. Disponer de los algoritmos que posibiliten la construcción automática de las URLs correspondientes a los registros en los catálogos de origen.

2. Que el programa de generación del catálogo colectivo sea modificado para realizar esa construcción automática de las URLs, las asocie con los acrónimos y construya las páginas html/xml incorporando esa URL.

1. Disponer de los algoritmos que posibiliten la construcción automática de las URLs correspondientes a los registros en los catálogos de origen.

Hay que examinar si las diferentes aplicaciones posibilitan ser "atacadas" mediante una URL directa que obvie el paso por el sistema jerarquizado de consulta de cada catálogo, así como el uso de la interfaz del catálogo en cuestión de que se trate.

Y hay que dilucidar si el algoritmo debe llevar unívocamente al registro "fuente" en la institución de origen.

Alternativa 1: usar el Código de Sistema

Si se cree que es así, que el enlace debe llevar unívocamente al registro "fuente" en la institución de origen, el medio más seguro es que la URL de acceso utilice como clave dentro del algoritmo el CÓDIGO DE SISTEMA del registro en el catálogo fuente, ya que es el único dato que simultáneamente:

- Es obligatorio para todos los registros,
- Es diferente para cada registro en cada sistema
- Enlaza de manera unívoca con el registro de origen y, por lo tanto, con las localizaciones de la publicación de que trate el registro.

Pero exige que ese código se conserve y se vaya trasladando en el proceso de aceptación y deduplicación de registros de la generación del CC.

Sin embargo, es necesario considerar que los datos facilitados a la Asamblea de REBIUN por el grupo de CC en sus informes muestran que en los catálogos fuente hay muchísimos registros que, aplicando las reglas del CC, son duplicados sin siquiera salir del catálogo de cada biblioteca. Es decir, el proceso de deduplicación del CC detecta (y resume en un solo registro contenedor) muchos registros duplicados internamente en los catálogos de cada biblioteca.

Eso significa que un registro contenedor del CC puede corresponderse con más de un registro en un mismo catálogo fuente. Y si es así, en esos casos, la elección del que va a ser considerado registro fuente se convierte en un problema. Y no sólo eso, sino que, además, a la hora de localizar ejemplares (origen de nuestro estudio) nunca se llegaría a los otros ejemplares existentes cuyas localizaciones estuviesen en los otros registros fuente equivalentes.

Alternativa 2: construir el algoritmo a partir de otros datos

Teniendo que localizar una construcción de una URL directa de un modo alternativo, al examinar los datos que podrían utilizarse, no se considera útil acudir a códigos normalizados como el ISBN o el ISSN porque son muchos los registros que carecen de ellos, sea por error, por ser la publicación lo suficientemente antigua, o porque se trate de un documento que no está sujeto a esas codificaciones. Se necesita, en todo caso, que se trate de un dato que esté siempre presente en los registros bibliográficos.

Se ha considerado que el campo de título es el idóneo, ya que su presencia está garantizada en los registros a incorporar al CC porque su ausencia es motivo de no aceptación del registro en el catálogo colectivo. Y porque cualquiera que sea el registro del CC todas las localizaciones que lleve aparejadas responden al mismo título.

En efecto, analizando el proceso que se realiza para la elaboración del CC en la sede de la empresa proveedora del servicio de CC a REBIUN, se observa que para formar la clave de eliminación de duplicados, se genera una primera "clave de 24 caracteres a partir de los datos del campo 245 \$a \$b. Estos datos se transforman en mayúsculas, sin tener en cuenta diacríticos ni signos de puntuación", se señala que a esa clave se le añaden otras,

y que, finalmente, son las claves totales resultantes las que se comparan, siendo que "en el momento en que la clave sea diferente, los registros se consideran diferentes."¹

¿Qué problemas puede plantear utilizar el campo título?

Puesto que en el proceso de depuración de duplicados queda garantizado que dos registros fuente con campo 245 diferente van a registro contenedor diferente, queda igualmente garantizado que la consulta del catálogo fuente por el título del registro contenedor necesariamente remitirá al registro fuente.

Sin embargo, junto a él pueden recuperarse más registros en el catálogo fuente, a saber, todos aquellos que cumplan con la regla de consulta, sea porque tienen las mismas palabras en diferente orden, sea porque tienen esas mismas y más palabras en el título, o sea porque efectivamente está duplicado el registro en el catálogo fuente (diferentes ejemplares de la misma obra catalogados por separado).

Eso significa que se accedería en el catálogo fuente a un punto de recuperación con más de un registro, requiriendo un paso más para acceder a la información sobre localización de ejemplares.

No obstante, eso no es necesariamente un inconveniente, ya que permite ayudar a superar el problema de los duplicados existentes en los catálogos fuente (lo que se mostraría serían esos duplicados –aunque puede que más obras-). Y porque en buena parte de los sistemas, cuando sólo existe un registro que cumpla los requisitos de la consulta, suele verse directamente el registro completo con las localizaciones, con lo que en esos casos no se daría el quedarse en la visualización previa o abreviada.

Sin embargo, cabe que esta "amplitud" provoque que, en determinados casos, la información sea excesivamente vaga (piénsese en un título que sea *Historia de España*, por ejemplo. Al objeto de que el resultado de la URL construida no devenga en un conjunto demasiado amplio de resultados, la mayoría de las cuales se refiera a obras que nada tienen que ver, sería de gran utilidad que la URL pudiera construirse lanzando una búsqueda en la que las palabras del título siempre fuesen consideradas con el operador "Y (AND)", en el orden el que se presentan, y/o combinadas con las palabras que compongan el autor, con las mismas consideraciones. Estas son, sin embargo, especificaciones de búsqueda que los propios sistemas de gestión de las bibliotecas pueden no tener.

De este modo parece conveniente concluir que es necesario indagar sobre las posibilidades de los diferentes sistemas para ser "consultados" directamente mediante una URL, centrándonos en las posibilidades de hacerlo apelando al CODIGO DE SISTEMA y a las PALABRAS DEL TITULO, ya que poder mejorar esa búsqueda estableciendo criterios algebraicos, de ordenación o de combinación no son posibles sin alguno de los anteriores.

Algoritmos de acceso directo en las diferentes aplicaciones de gestión de catálogos en las Bibliotecas de REBIUN.

1.- Por Código de registro en el sistema de origen:

De la búsqueda realizada, aparentemente, sólo dos sistemas, ALEPH y SABINI, posibilitan el acceso acudiendo al número de sistema. Sin embargo, aunque en las instalaciones ABSYS no parezca utilizarse, lo cierto es que la manera en la que actualmente facilita el CC de REBIUN el acceso a los registros originales de bibliotecas con ABSYS, es precisamente a través del código de registro en la biblioteca original (al menos desde la puesta en marcha de la nueva versión en junio 2004). Este hecho hace pensar que el código de identificación de cada registro en el sistema de origen se conserva en el proceso de eliminación de duplicados del Catálogo colectivo, ya que, de lo contrario, no sería posible acceder al registro origen en una instalación ABSYS de una obra de la que el registro contenedor en el CC no fuera de esa instalación –como sí está sucediendo-.

¹ Del informe del Grupo de CC a la Asamblea de REBIUN en Noviembre de 2002

Caso ABSYS:

`http://<v1>/cgi-bin/abweb/L1/T1?ACC=DIR&EXP=<v2>.TITN.`

donde <v1> es el valor a tomar por el servidor
donde <v2> es el código del registro en el sistema

valores de <v1>
Servidor UJA = `www2.ujaen.es`

Caso ALEPH:

`http://<v1>/F?func=find-b&find_code=SYS&request=<v2>`

donde <v1> es el valor a tomar por el servidor
donde <v2> es el código del registro en el sistema

valores de <v1>
Servidor CSIC = `aleph.csic.es`

Caso SABINI:

`http://<v1>?C4=N.MARC&C5=<v2>`

donde <v1> es el valor a tomar por el servidor
donde <v2> es el código del registro en el sistema

valores de <v1>
Servidor UPNA = `brocar.unavarra.es/cgi-bin/sb3spa`

2.- Generación de algoritmos a partir del título (§a del campo 245).

Caso ALEPH:

`http://<v1>/F?func=find-b&find_code=WTI&request=<v2>`

donde <v1> es el valor a tomar por el servidor
donde <v2> es la suma de las palabras del §a de 245, sustituyendo los signos de puntuación por su equivalente (una coma es %2c), por ejemplo, `EI+alcalde+de+zalamea`

valores de <v1>
Servidor de CSIC = `aleph.csic.es`

Caso AMICUS/LIBRIVISION

La mayoría de las instalaciones de AMICUS carecen de esta posibilidad al controlarse previamente la sesión, pero un desarrollo que se está terminando en la UPV contiene esa especificación entre sus resultados

Caso MILLENIUM

`http://<v1>/search*spl /t?SEARCH=<v2>`

donde v1 es el servidor
donde v2 es la suma de las palabras del §a de 245, los signos de puntuación se omiten y la ñ vale como ñ o como ~a

valores de <v1>
UNAV = www.unav.es

Caso SABINI

<http://<v1>?C4=T%EDtulos&C5=<v2>>

donde v1 es el servidor
donde v2 son las palabras del \$a de 245, sumadas o con sus espacios en blanco, con o sin diacríticos.

valores de <v1>
UPNA = brocar.unavarra.es/cgi-bin/sb1spa

Caso UNICORN

<http://<v1>/uhtbin/cgisirsi/x/0/5?searchdata1=<v2>{245}>

donde <v1> es el servidor
donde <v2> es la suma de las palabras del \$a de 245, sustituyendo los blancos por el símbolo +, los signos de puntuación se mantienen, Ojo con la ñ.

Valores de <v1>
UAM = oruga.bibcen.uam.es

Caso VTLS:

<http://<v1>/vtls.web.gateway?searchtype=title&conf=080000+++++++&searcharg=<v2>>

donde <v1> es el valor a tomar por el servidor
donde <v2> es la suma de las palabras del \$a de 245, sustituyendo los signos de puntuación por su equivalente (una coma es %2c), por ejemplo, El+alcalde+de+zalamea

Valores de <v1>
Servidor de BC = www.gencat.net/cgi-bin/bc/biblio

Caso Universidad de VIC

La Universidad de Vic dispone de un MicroVTLS sobre el que han dispuesto una pasarela para posibilitar su consulta vía web. Se ha comprobado el funcionamiento del siguiente algoritmo :

<http://<1>/cgi-bin/biblioteca/cerca.pl?quant=0&seleccio=Titol&nom=<v2>>

donde <v1> es el valor a tomar por el servidor
donde <v2> es la suma de las palabras del \$a de 245

valores de <v1>
Servidor de UVIC = www.uvic.es

Caso universidades con sistema Biblio3000

En estos momentos, en el sistema Biblio3000 no se facilita la posibilidad de atacar mediante ninguna URL directa ningún punto del catálogo diferente de la propia consulta. No obstante, al habla con la empresa productora, nos han informado de que en el momento en el que

REBIUN tomara una decisión desarrollarían la programación necesaria para facilitar ese acceso desde el CC.

Otros sistemas

Algunas bibliotecas universitarias disponen de otros sistemas (Liber-Marc, el gestor Loris, desarrollos propios bajo Php, etc.), que no existen más que en una o dos bibliotecas cada uno. En total, no llegan a 8 sobre un total de 68 bibliotecas, por lo que en esta fase de propuesta se ha obviado su detalle y análisis particular.

2.- Que el programa de generación del catálogo colectivo sea modificado para realizar esa construcción automática de las URLs, las asocie con los acrónimos y construya las páginas html/xml incorporando esa URL.

A la vista de las realidades de los algoritmos de cada aplicación, y de las consecuencias en los resultados de los accesos según qué tipo de algoritmo se emplee, da la impresión de que la solución más sensata y rápida, de menor trabajo para el mayor número de instalaciones, es la de que sea posible construir el algoritmo a partir de las palabras del título o a partir del número de registro.

Como en ambos casos se trataría de la construcción de un algoritmo para el que los datos necesarios de la obra en cuestión están disponibles en el registro contenedor del CC del que se quieren conocer las localizaciones, caso del título, o en la información mantenida en el procedimiento de eliminación de duplicados y asignación de localizaciones

como las direcciones de los servidores de las diferentes bibliotecas han de considerarse como estables (el cambio se produce en pocas ocasiones, normalmente en intervalos mayores de 5 años),

como es más acorde con la filosofías base de los programas actuales no acumular información que pueda ser construida cuando se requiera un nuevo pinchazo en la pantalla para que las páginas sean más ligeras y ágiles

concluimos que resultaría idóneo

1. recopilar el 100% de los valores <v1> a tomar por los servidores de cada universidad
2. recopilar "oficialmente" si los sistemas permiten acceder a través del código de registro en su sistema, a través de las palabras del \$a del campo 245 y/o a través de una combinación de las palabras del \$a del campo 245 y (en su caso) de las palabras del \$a de los campos 1XX.
3. facilitar a la empresa proveedora del servicio de CC la relación tanto de estos valores como del algoritmo de generación de la URL para cada caso
4. que la empresa desarrolle la mínima programación (porque es mínima) para que, cuando se "pinche" un acrónimo de una biblioteca en un registro del catálogo, se ejecute el programa que tome el número de registro o las palabras del \$a del campo 245 y (en su caso del \$a de los campos 1XX) y construya, en función del acrónimo, la URL correspondiente y ejecute su visualización en una ventana nueva.
5. Que en el caso de que ABSYS no realice esos cambios, tampoco facilite vía CC acceso a los registros en el caso de instalaciones ABSYS.

Con esto quedaría resuelto el acceso a las localizaciones de cada universidad ligadas a un registro del catálogo colectivo.