

CATALOGACIÓN COOPERATIVA DE INTERNET: HERRAMIENTA PARA DESARROLLAR COLECCIONES DE CALIDAD

*Manuel Antonio Diosdado Barrón**

Introducción

El crecimiento del World Wide Web y la naturaleza dinámica de los recursos disponibles en Internet representan un reto para los organizadores de la información y han hecho necesaria la creación de herramientas que proporcionen acceso a los millones de documentos que existen en formato electrónico. En efecto, una página web puede ser movida de lugar dentro del mismo sitio, o desaparecer por completo. Su contenido puede ser modificado con gran rapidez, puede cambiar la persona responsable de la página, o la institución auspiciadora. Todo esto sin contar que se pueden agregar nuevos métodos de acceso diferentes a los originales. Por medio de las herramientas de búsqueda existentes se ha pretendido dar solución a estas variables. Estas herramientas utilizan métodos automáticos para identificar los

* Bibliotecario, Biblioteca Daniel Cosío Villegas, El Colegio de México.

recursos de Internet. De hecho, existen muchos programas que "navegan" automáticamente a través de la Web, buscando enlaces, recuperando documentos, indizándolos y creando bases de datos. Estos sistemas recuperan gran cantidad de documentos, pero su precisión es muy baja. La causa principal no es que los métodos automáticos utilizados describan inadecuadamente los recursos electrónicos de Internet, sino que los mismos documentos HTML carecen de información suficiente que describa el recurso adecuadamente. Esto arroja como resultado mucha "basura" en la recuperación de información. Internet, se dice, crece de una manera exponencial, se asume que será cada vez más difícil navegar y, por lo tanto, localizar información relevante.

Conscientes del crecimiento de Internet y de las desventajas que ofrecen los procedimientos existentes de descripción de su información y sus métodos, bibliotecarios, informáticos y diseñadores de software, entre otros profesionales, ya han planteado la necesidad de crear descripciones y catálogos que logren identificar los recursos electrónicos en Internet de una manera eficaz que permita una búsqueda y localización más efectiva. Así, debemos reconsiderar la manera en que este tipo de recursos pueden ser representados y organizados de una mejor forma.

La organización de información en INTERNET

La necesidad de organizar la información de Internet ha llevado a la comunidad académica a emprender diversos esfuerzos por controlarla. Entre estos esfuerzos destacan principalmente dos: los motores de búsqueda, por un lado, que son utilizados como una herramienta de proyectos patrocinados por instituciones que tienen como objetivo agrupar la información de recursos en Internet y, por otro lado, el que ha resultado del "enfoque metadata". Los primeros, como ya se sabe, son "una herramienta que permite acceder mediante palabras clave, los recursos de Internet indizados en su base de datos" (Pedraza Gracia 1997). Previa a esta indización, se realiza una selección y una evaluación de cada uno de los recursos, con criterios establecidos por cada proyecto. Por lo que toca al enfoque metadata, en el cual nos centraremos, este término "se utiliza para designar el conjunto de

elementos informativos que pueden emplearse para describir y representar objetos de información" (Pedraza Gracia 1997). Entre las propuestas derivadas de este concepto tenemos el URC (Características de Recurso normalizado), TEI (Iniciativa para la codificación de textos), el DC (Núcleo Dublín). Encontrar una solución al problema de la organización bibliográfica en Internet se ha convertido en materia de estudio de muchas instituciones, entre las que destaca los esfuerzos del Online Computer Library Center (OCLC). Ya en 1996 Hsieh-Yee puntualizaba que aun cuando era deseable contar con un sistema de información que organice todo lo que existe en Internet, en ese momento tal sistema no era factible ni podría ser efectivo por las siguientes razones:

1. Existen demasiados recursos en Internet
2. La calidad de muchos de esos recursos es cuestionable
3. Muchos recursos son de naturaleza efímera y pueden ser de escaso valor para los usuarios (Hsieh-Yee 1996)

Entre los primeros intentos, encaminados a dar un orden a Internet, figuran The Argus Clearinghouse. Este proporciona un acceso central temático, el cual identifica, describe y evalúa información de recursos en Internet. Esta actividad es regida por una serie de criterios como son: el nivel de descripción del recurso, el diseño, la organización de sus índices, así como la metainformación que proporcione¹. Otro más es CyberStacks, el cual se define a sí mismo como "Colección centralizada, integrada y unificada de recursos Internet". Este es un sitio experimental que ordena y selecciona información de recursos Internet en las áreas de ciencia y tecnología, de acuerdo a la clasificación de la Biblioteca del Congreso de los Estados Unidos, utilizando criterios tales como autoridad, exactitud, claridad, originalidad, etc. Otro proyecto, el Infofilter, presenta una serie de criterios para seleccionar y evaluar recursos en Internet, como son, autoridad,

¹ Los criterios se pueden examinar en la página web de Argus Clearinghouse: <http://www.public.iastate.edu/~CYBERSTACK/signif.html>. Véase anexo para las páginas web de otras iniciativas

contenido, organización, aceptación o uso general, motores de búsqueda y diseño de gráficos entre otros. El sitio se mantiene, pero el proyecto cesó en julio de 1997. Otro esfuerzo lo representa la Internet Public Library, una iniciativa de la University of Michigan, que también ha definido una serie de políticas de selección y evaluación, como son contenido, actualidad, disponibilidad, etc. McKinley Group, subsidiaria de Excite, Inc. desarrolla a Magellan Internet Guide, un directorio de recursos organizado en clases o categorías temáticas que para incluir algún recurso de Internet ha fijado criterios como son profundidad y facilidad de exploración. Un esfuerzo más es la World Wide Web Virtual Library, la cual agrupa recursos electrónicos de Internet que, de igual forma, son sometidos a ciertos criterios de selección.

Es importante puntualizar que las iniciativas con enfoque metadata se plantean, como objetivo principal, el que cada uno de los recursos sea "catalogado", para lo cual proponen un cierto número de elementos que describirían al recurso, y que entonces, se volverían susceptibles de ser llaves de localización. El enfoque de los motores de búsqueda deja a un lado la propuesta de organización en su estructura de metadatos. Su principio de agrupamiento se basa en establecer una serie de elementos que deben reunir cada uno de los recursos seleccionados de Internet con el objetivo de evaluar su posible integración a su base de datos. Esta característica garantiza la calidad de la información. Estas iniciativas son limitadas porque algunas sólo enfocan algunas áreas en específico, además de que no se hace una convocatoria pública que permita la participación de otras instituciones interesadas.

OCLC y la organización de INTERNET: el Proyecto CORC

Consideramos que una combinación de ambos enfoques la encontramos actualmente en una iniciativa de OCLC. Esta pretende crear una base de datos con recursos de Internet, los cuales se describen con un enfoque de metadatos, ya sea usando el formato MARC21 o el Núcleo Dublin. Además, se pretende que toda institución que participe aportando recursos establezca un adecuado cuerpo de criterios para la selección y evaluación de los mismos.

OCLC puso a disposición su sistema automatizado para recursos electrónicos, en una primera etapa de prueba, hasta junio de 2000, la cual fue de acceso gratuito y a partir de julio del 2000 ya como un servicio formalizado, pretende sumar los esfuerzos cooperativos de las bibliotecas interesadas, encaminado a la organización de los recursos electrónicos de Internet. No se incluyen todos los recursos, sino sólo aquéllos que han sido evaluados por alguna biblioteca y que sean de probada calidad. Este proyecto es denominado Cooperative Online Resource Catalog (CORC).

El proyecto CORC es definido por OCLC como "un sistema de creación de metadatos basado en el Web, optimizado para crear registros bibliográficos y rastreadores (Pathfinders) de recursos electrónicos, tanto a nivel local como aquellos basados en el Web". CORC se conforma de cuatro bases de datos que apoyan la creación de registros:

- CORC Resource Record Database
- CORC Authority Database
- CORC Pathfinder Database
- Dewey Decimal Classification Database

CORC ofrece tecnología avanzada para apoyar los siguientes procesos automatizados: a. Creación de registros, b. Control de autoridades, c. Mantenimiento de URL, d. Asignación de encabezamiento de materia, e. Creación de rastreadores (pathfinder).

OCLC extendió la invitación, en su fase experimental, a todas las bibliotecas interesadas en participar y cooperar, para construir una base de datos de calidad con archivos de recursos de la Web seleccionados por las bibliotecas. Actualmente participan más de 350 instituciones, principalmente de los Estados Unidos y Canadá. Para detalles de las características del servicio de CORC, véase su página web.²

² Características del servicio CORC: <http://www.oclc.org/oclc/corc/>

Participación de la BDCV en CORC

La Biblioteca Daniel Cosío Villegas (BDCV) de El Colegio de México, institución docente y de investigación, consciente de las necesidades de sus usuarios en la era del conocimiento digital, así como del aprovechamiento de las nuevas tecnologías en lo que concierne a la recuperación de recursos electrónicos y del crecimiento de Internet, ha tomado la decisión de sumarse a los esfuerzos cooperativos de otras bibliotecas en el sistema CORC, junto con la Universidad Autónoma de Ciudad Juárez, Chihuahua y el ITEMS, todas de México.

Agregarse al esfuerzo conjunto implica puntualizar las principales actividades a realizar con el uso de CORC en la BDCV:

- a. Cotejar los recursos electrónicos que actualmente se listan en la página Web de la BDCV, integrados por alrededor de 500 registros, con la base de registros de CORC.
- b. Exportar de CORC a nuestra base de datos (ALEPH) los registros identificados como existente.
- c. Modificar la catalogación de acuerdo a la normatividad y políticas locales de la BDCV.
- d. Crear en CORC la catalogación de los recursos que no se localicen.
- e. Exportar los registros creados en CORC a nuestra base de datos (ALEPH).
- f. Modificar la catalogación de los registros creados de acuerdo a la normatividad y políticas locales de la BDCV.
- g. Buscar e identificar rastreadores (pathfinder) en la base de CORC, que sean pertinentes a las temáticas de investigación y docencia de El Colegio, además de ser evaluados de acuerdo a los criterios de selección de recursos electrónicos de la BDCV.
- h. Modificar el rastreador identificado, de acuerdo a las indicaciones resultantes de la evaluación.
- i. Exportar de CORC los rastreadores (pathfinder) seleccionados y/o modificados a nuestra base de datos (ALEPH).
- j. Incorporar los rastreadores exportados a la página Web de la BDCV.

- k. Buscar e identificar registros en CORC, pertinentes a las temáticas de investigación y docencia de El Colegio, con el fin de agruparlos temáticamente.
- l. Construir en CORC rastreadores (pathfinder), con los registros seleccionados y agrupados temáticamente.
- m. Exportar de CORC los rastreadores construidos a nuestra base de datos (ALEPH)
- n. Incorporar los rastreadores exportados a la página Web de la BDCV.
- o. Búsqueda en CORC de registros de autoridad.
- p. Exportar los registros de autoridad de CORC (personales, corporativos, temáticos), a nuestra base de datos (ALEPH).
- q. Modificar la estructura de los registros de autoridad de acuerdo a la normatividad y políticas locales de la BDCV.

Las actividades descritas se han integrado ya a los procesos de trabajo cotidianos de la BDCV, de acuerdo a su naturaleza, en las siguientes fases:

Fase I:

Paso 1. Se exportarán de CORC a la base local (ALEPH), todos los registros identificados, tomando como base los listados de los recursos electrónicos por Centro de estudios.³

Paso 2. Ya en la base local, se procederá a la importación, conversión (utilizando la norma 2709 OCLC) y edición del registro.

Paso 3. Con el registro editado se realizan las modificaciones necesarias de acuerdo a la normatividad y políticas de la BDCV, que básicamente son:

- a. Adaptación y adición de encabezamientos de materia.
- b. Asignación de la clasificación local de la BDCV (RE/n...)
- c. Modificación y validación de la catalogación descriptiva del recurso. En este tipo de recursos son muy importante las

³ En la página Web de la BDCV, están listados los recursos electrónicos seleccionados por Centros de estudios así como de algunos Programas de El Colegio.

notas, principalmente la de resumen (520/MARC/21) la cual debe ser adecuadamente traducida.

d. Salvar el registro en la base de datos local; el sistema le asignará un número.

Paso 4. Se repetirá este procedimiento hasta agotar los recursos identificados en la base de registros de CORC.

Fase II:

Paso 1. Proceder a la creación de la catalogación de registros de recursos en CORC.

Paso 2. Se procede a su validación en CORC, si hay error el sistema lo informará, para realizar las correcciones necesarias.

Paso 3. Se procede a guardar el registro en una base temporal de CORC.

Paso 4. Cuando ya el registro está completo se agrega a WorldCat, base principal de OCLC; el sistema le asigna un número.

Paso 5. Es exportado a un sistema local, con la rutina de la fase I.

Fase III:

Paso 1. Se recupera cada uno de los rastreadores identificados previamente.

Paso 2. Se hace una copia del rastreador (se clona).

Paso 3. Se recuperan uno por uno los rastreadores localizados en CORC, con el fin de editarlos y modificarlos, personalizándolos de acuerdo a la institución.

Paso 4. Se exportan de CORC los rastreadores a nuestro sistema local.

Paso 5. Se procede a su incorporación a la página web.

Paso 6. Posteriormente se exportarán todos y cada uno de los registros que conforman el rastreador, procediendo como se indica en la Fase I.

Fase IV:

Paso 1. Se buscan e identifican registros en la base CORC, de acuerdo a los intereses de investigación de El Colegio de México.

Paso 2. Se agrupan temáticamente.

Paso 3. Se procede a la construcción en CORC de rastreadores con los registros seleccionados.

Paso 4. Se editan los rastreadores para personalizarlos de acuerdo a la institución.

Paso 5. Se validan y se graban en el sistema CORC.

Paso 6. Se hace una copia del rastreador construido y se procede a su exportación a nuestro sistema local.

Paso 7. Se procede a su incorporación a la página Web de la BDCV.

Fase V.

Paso 1. Se buscan registros de autoridad en CORC.

Paso 2. Se procede a su exportación a nuestro sistema local.

Paso 3. Ya en el sistema local, se procederá a la importación, conversión (utilizando la norma 2709 OCLC) y edición.

Paso 4. Se graban en el sistema local; el sistema asigna un número.

Estas son las fases que se han propuesto para la catalogación e incorporación de los recursos electrónicos que se tienen seleccionados en la BDCV y los que a corto plazo se deben buscar en el universo de CORC (hasta septiembre de 2000 sobrepasan los 350,000 registros), así como la incorporación de este tipo de recursos informativos de parte de la BDCV como biblioteca cooperante del sistema CORC. Se ha trabajado de manera piloto el total de las fases descritas. Por el momento, sólo se ha instrumentado la Fase I de esta propuesta, en la que parte del personal de procesos técnicos ya está trabajando.

Resultados preliminares del Proyecto.

Se calcula, de acuerdo a las pruebas piloto realizadas, que aumentará eficientemente la catalogación de estos recursos, y que se simplificará la creación y mantenimiento de rastreadores. Esto impactará de forma favorable el suministro de este servicio a los

usuarios ya que es posible integrar en un rastreador recursos físicos y digitales. Los recursos físicos incluyen la información almacenada en formatos convencionales como son: papel, microformatos, audiovisuales, etc., que al ser conjuntados en un rastreador con recursos digitales, y descritos con mayor detalle, se hacen más atractivos para los usuarios. En este punto es importante señalar la verificación de enlaces que realiza también el sistema CORC, el cual notifica a cada biblioteca cuando un URL es interrumpido o desviado. Una sola actualización beneficia inmediatamente a todas las bibliotecas que hayan realizado el enlace del URL. Entonces, a largo plazo, la BDCV tendría una gran cantidad de recursos electrónicos de calidad, catalogados en su sistema, lo cual permitirá al usuario, antes de decidirse a consultarlo, tener a la mano información sobre el recurso, desde el título de la página, la institución que la patrocina, hasta detalles como tipo de información que la constituye, contactos, materias de que trata el contenido, desde que año está activo, modo de acceso, etc. Además, el usuario encontraría en la página Web de la BDCV una gran cantidad de rastreadores (bibliografías anotadas de recursos electrónicos y físicos) por áreas de sus interés.

El futuro de esta modalidad de información se puede ver en tres escenarios, uno inmediato, el cual esta conformado por la comprensión y adopción de los mecanismos que maneja el sistema CORC de OCLC, así como los del sistema local, adopción encaminada a ejecutar la instrumentación de los procesos conjuntos de los sistemas, así como a la rutinización de estos procesos en las labores cotidianas de catalogación del centro de documentación participante.

Un segundo escenario se vislumbra en el aprovechamiento racional que se haga de los recursos con que cuenta la base de CORC, en apoyo a un eficiente desarrollo de la colección de la biblioteca, así como la aportación de cada una de las bibliotecas participantes a la base CORC, con recursos originales, para así cerrar el círculo de cooperación.

Finalmente, una tercera etapa se integraría como en toda biblioteca, el mantenimiento de la colección, implicando su actualización y descarte, acciones en las cuales también es fundamental el sistema CORC.

Conclusiones

El desorden creciente de Internet no se puede detener, pero si podemos seleccionar parte de ese universo para incorporarlo a nuestros catálogos en OPAC's y páginas Web. Las tecnologías de informática y las telecomunicaciones están transformando la forma en que el conocimiento se produce, se almacena, la forma de acceder a él, su disseminación y su preservación. Por todo esto, debemos conformar un *hábeas* de criterios que nos permitan garantizar un mínimo de calidad en el desarrollo de la colección de recursos electrónicos. El universo de selección ha aumentado a tal grado que las instituciones enfrentan dificultad para rescatar información valiosa. Las propuestas de enfoque metadata, así como las iniciativas que buscan agrupar en razón de la calidad, son herramientas muy importantes que debemos utilizar combinando la experiencia de los profesionales de la biblioteca, la fuerza de la cooperación y los beneficios de la tecnología. Consideramos que CORC, al hacer extensivo su proyecto a toda institución interesada en apoyar y beneficiarse de él, está retomando el concepto fundamental de la cooperación que, en última instancia, servirán al usuario de nuestras bibliotecas

BIBLIOGRAFÍA

- Butterfiel, Kevin L. *Cataloger's and the creation of metadata systems: a collaborative vision at the University of Michigan*. OCLC Internet Cataloging Project Colloquium. Ann Arbor, Mich : University of Michigan, octubre 1998. URL: <http://www.oclc.org/oclc/man/colloq/butter.htm>
- Dublin Core metadata initiative: <http://purl.org/DC/documents/rec-dces-19990702.htm>, 1999.
- Hsieh-Yee, Ingrid. *Modificaciones a las reglas de catalogación y la infraestructura de OCLC para una organización efectiva de los recursos de Internet*. URL:http://cuib.laborales.unam.mx/-felipe/cat_Internet/modificaciones.htm

Library of Congress. Network Development and MARC Standards Office. *Guidelines for the use of field 856 (revised August 1997)*

URL: <http://lcweb.loc.gov/marc/856guide.htm>

Olson, Nancy E., ed. *Cataloging Internet resources : a manual and practical guide*. University of Oregon, Library OCLC's featured members, OCLC Online Computer Library Center, October 1998.

URL: <http://www.oclc.org/oclc/man/9256cat/toc.htm>

Pedraza Gracia, Manuel José. "Aproximaciones a la catalogación de documentos electrónicos de acceso remoto". *Scire: representación y organización del conocimiento*. vol. 3, no. 1, enero-junio, 1997. pp 99-125.

ANEXO

The Argus Clearinghouse

URL:<http://www.clearinghouse.net/>

Criterios: URL:<http://www.clearinghouse.net/ratings.html>.

Biblioteca Daniel Cosío Villegas, El Colegio de México

<http://200.12.162.127/>

CyberStacks

URL:<http://www.public.iastate.edu/~CYBERSTACKS/>

Criterios: URL:<http://www.public.iastate.edu/~CYBERSTACKS/signif.htm>

Infofilter

<http://www.usc.edu/users/help/flick/Infofilter/>

Criterios: <http://www.usc.edu/users/help/flick/Infofilter/template.html>

Internet Public Library

<http://www.ipl.org/>

Criterios: <http://www.ipl.org/ref/RR/Rabt.html#policy>

Magellan Internet Guide

<http://www.mckinley.com/>

Criterios: <http://www.lib.ua.edu/maghelp.htm#howdoes>

Núcleo Dublín

<http://purl.org/DC/documents/rec-dces-19990702.htm>

TEI (Iniciativa de Codificación de Texto)

<http://www.uic.edu/orgs/tei/index.html>

World Wide Web Virtual Library

<http://www.ciolek.com>

Criterios: <http://www.ciolek.com/WWWVLPages/QltyPages/QltyLinks.html>