

V Coloquio de Tecnologías Aplicadas a los Servicios de Información

Marc, Metadata y Z39.50

Mayo, 2000

**Guilda Ascencio
Multimedium Sistemas de Información
Tif. 7612965
E-mail: ascencio@cantv.net
msi@reacciun.ve
Caracas, Venezuela**

**María Eugenia Franceschi
Centro de Información Técnica (CIT)
INTEVEP
Tif. 9086896
E-mail: mfrances@intevp.pdv.com
Caracas, Venezuela**

Resumen

Tres estándares mantienen focalizada la atención de la comunidad bibliotecaria internacional: **Marc, Metadata y Z39.50**

El formato **MARC** (machine readable catalog = catálogo legible por máquina) es actualmente una familia de formatos, todos con una estructura de registro similar, pero con diferencias significativas en la manera de implementación. Originado a finales de 1960 como respuesta a las necesidades de computarización de las bibliotecas, ha sido el más duradero y desarrollado de los formatos de metadada. Aún hoy día es el medio por excelencia para el intercambio de la catalogación de registros entre bibliotecas.

Actualmente es el formato que han adoptado casi todos los productores de software de automatización de bibliotecas, como estructura básica para manipular los registros bibliográficos para su despliegue e indización.

El uso cada vez mayor de Internet exige el desarrollo de herramientas de navegación efectivas y fáciles de utilizar que permitan obtener una información de calidad. Una parte esencial del acceso a los recursos de Internet es la existencia de un índice que economice el tiempo que los usuarios dedican a las búsquedas de información. Por ello, bibliotecarios y otros profesionales de la información discuten actualmente métodos para el acceso a los recursos de la Red Internet, teniendo como meta la creación de un catálogo de documentos Internet que se asemeja y sea compatible con las bases de datos locales.

Dublin Core Metadata Element Set, es uno de los estándares de metadata que tiene por objeto facilitar la recuperación de recursos digitales ubicados en Web, al proporcionar un grupo mínimo de elementos descriptivos, compatibles con el formato MARC, los cuales pueden ser incluidos en los documentos electrónicos.

El protocolo **Z39.50** complementa el concepto de acceso global a la información. Z39.50 dirige la comunicación entre aplicaciones de recuperación de información, cliente y servidor, las cuales pueden residir en diferentes computadores.

El protocolo se propuso originalmente en 1984 para el uso con información bibliográfica. A medida que el interés sobre el mismo se amplió, se estableció en 1990 el Grupo de Implementadores del Z39.50 (ZIG). Los miembros incluyen: manufactureros, vendedores, consultores, proveedores de información y universidades, quienes desean acceder o proveer acceso a varios tipos de información, incluyendo bibliográfica, texto, imagen, financiera, de utilidad pública, química y noticias

Palabras claves: MARC, METADATA, Z39.50, ACCESO A BASES DE DATOS, BASES DE DATOS

Currículo

Guilda Ascencio

Especialista en sistemas de información, trabaja en estrecha relación con las bibliotecas y centros de información de la región desde 1987. Desde esa fecha participa en forma independiente ó con la Unesco, en actividades tendientes a la difusión y capacitación de usuarios en la plataforma CDS/Isis y su familia de productos, así como en asesorías a diferentes instituciones nacionales y regionales. Responsable del desarrollo del programa **Documentum**, Sistema de Gestión de Información Documental para bases de datos mantenidas bajo al estructura de archivos CDS/Isis.

María Eugenia Franceschi

Traductora egresada de la UCV en los idiomas español, francés e inglés, con maestría en ciencias del lenguaje de la Université de la Sorbonne Nouvelle, París, Francia. Desde 1985 trabaja para el Centro de Información Técnica de PDVSA- Intevep, donde ha desempeñado diversas funciones, entre las cuales editora técnica, responsable del desarrollo y mantenimiento del tesoro de la industria petrolera para la base de datos de la Red de Información Petrolera y Petroquímica –Rippet-, administradora de la base de datos corporativa RIPPET y responsable de la creación y mantenimiento de la base de datos corporativa en la Intranet de PDVSA.

No cabe duda de que Internet ha revolucionado la manera de interactuar con el mundo.....

Cuántos de nosotros no hemos experimentado una gran desilusión al tratar de “navegar” por Internet usando algunos de los motores de búsqueda, como Altavista, HotBot o Infoseek, entre otros, con el objeto de buscar algo en particular, y obtener como resultado cientos de miles de documentos, y no exactamente lo que buscamos. Gran parte del material útil que se encuentra en la Web es difícil de localizar y por lo tanto, a los efectos, no está disponible. De allí que esta red de redes se haya convertido en un arma de doble filo; por cuanto es cada vez más fácil publicar cualquier tipo de información, pero resulta cada vez más difícil localizar información relevante.

De acuerdo con un artículo publicado por un grupo de trabajo australiano relacionado con la navegación de la información gubernamental [1], existe una serie de inconvenientes con los motores de búsqueda de Internet:

- *la información relevante puede perderse por cuanto las páginas contienen tipos de recursos adicionales a HTML (por ej.: imágenes, bases de datos, documentos PDF);*
- *los motores de búsqueda por lo general no recogen todas las páginas de un lugar, sino sólo los dos o tres principales niveles jerárquicos, con lo que se omiten documentos importantes, los cuales en lugares más complejos pueden ser localizados en niveles de jerarquía inferiores;*
- *los motores de búsqueda, en especial los más completos, pueden indizar sitios sobre una base no frecuente y por lo tanto pueden no contener la información más actualizada; y*
- *es posible localizar información irrelevante, por cuanto los motores de búsqueda no tienen forma (o muy poca) de distinguir entre palabras importantes y fortuitas en el texto del documento.*

Ante este panorama, encontramos el término “metadato” o como usualmente se le denomina “la metadata”. Hoy día resulta difícil encontrar una publicación sobre recuperación de recursos electrónicos que no hable sobre este aspecto, si bien algunos desconocen exactamente de qué se trata. Para muchos es algo completamente novedoso e innovador; en el caso de los profesionales de la información, como los catalogadores e indizadores, puede que igualmente lo sea, pero en realidad han estado trabajando con metadata por años si saberlo. Así lo indican Jessica Milstead y Susan Feldman [2], “como el hombre

que había estado escribiendo prosa durante toda su vida sin saberlo, los bibliotecólogos e indizadores han estado produciendo y estandarizando metadata por siglos. Ignorando esta herencia, una gran variedad de otros interlocutores han entrado recientemente en el juego, y muchos de ellos no tienen idea de que ya alguien más “ha estado allí, y hecho eso”.

Ahora bien, ¿qué es metadata? Si analizamos la composición de esta palabra, observaremos que su primer elemento “meta” proviene del griego con el significado de además, más allá, después, y “data” del inglés con el significado de información real. Así pues, en simples términos, se trata de “datos sobre datos”. Son datos estructurados que describen las características de un recurso. Un registro catalográfico es metadata. Priscilla Caplan [3], por ejemplo, la ha definido como una forma más neutra de catalogar sin el “exceso de equipaje” de las reglas Anglo-americanas de catalogación. Este concepto refiere hoy día a cualquier dato usado para ayudar a la identificación, descripción y localización de recursos electrónicos.

Entonces, si retrocedemos en el tiempo, para los profesionales de la información el concepto de metadata es sinónimo de catalogación e indización. En otras palabras, información bibliográfica estándar, resúmenes, términos de indización son substitutos de metadata. Y en este ámbito el formato más expandido y utilizado es el formato MARC.

Pasemos ahora a analizar el formato MARC, de forma de comprender más fácilmente el por qué de la metadata en la actualidad.

MARC, acrónimo de *MAchine Readable Cataloguing* o lo que se conoce en español como catalogación legible por máquina, es un formato de codificación utilizado para el almacenamiento de datos bibliográficos legibles por máquinas. Lo anterior implica que una computadora puede leer e interpretar los datos de un registro catalográfico. Este formato data de los años 60, cuando la Biblioteca del Congreso Norteamericano —el repositorio oficial de las publicaciones norteamericanas- comenzó a utilizar computadoras e ingenió un sistema de números, letras y símbolos dentro de un registro catalográfico para marcar los diferentes tipos de información. El formato LC MARC original se convirtió en US MARC. A medida que los sistemas de bibliotecas se fueron computarizando, MARC se adoptó en el software de automatización como la estructura básica para manipular los registros bibliográficos y realizar su despliegue e indización.

Este formato comprende una estructura de datos y un procedimiento de codificación implementado en las normas ANSI Z39.2 (Information Interchange Format) e ISO 2709 (Format for Information Exchange). El mismo se utiliza tanto en América, primordialmente en Estados Unidos, como en Europa, para el almacenamiento y recuperación de información bibliográfica, al igual que para el intercambio de información con las mayores bases de datos de bibliotecas OCLC (On-line Computer Library Catalogue) y RLIN (Research Libraries Information Network).

En Estados Unidos se conoce como USMARC, cuyos responsables son la Biblioteca del Congreso Norteamericano, bajo la figura de MARC Standards Office, y MARBI -**MA**chine-**R**eadable **B**ibliographic Information Committee- de la Asociación Norteamericana de Bibliotecas (ALA). Asimismo existe el CAN/MARC, para Canadá, y el UKMARC, para el Reino Unido, bajo la responsabilidad del British Library. Finalmente, IFLA -International Federation of Library Associations and Institutes- desarrolló el formato UNIMARC, pues si bien los formatos MARC nacionales tienen mucho en común, son lo suficientemente diferentes como para requerir una conversión mediante el uso de programas. Por lo tanto UNIMARC fue diseñado con la idea de funcionar como un formato intermedio.

En 1994, los representantes de British Library -UKMARC-, de la Biblioteca del Congreso Norteamericano -USMARC- y de la Biblioteca Nacional del Canadá (BNC) -CAN/MARC- comenzaron un proceso de armonización de los diferentes formatos nacionales. Se espera que durante el presente año se implanten los acuerdos, los cuales permitirán mejorar el intercambio de datos, eliminar la necesidad de programas de conversión y probablemente reducir los costos relacionados con el mantenimiento del formato y con la documentación.

El uso del formato MARC facilita la catalogación al no duplicar el trabajo, así como permitir a las bibliotecas compartir recursos. Asimismo, el uso del formato MARC permite a las bibliotecas y centros de información hacer uso de los sistemas automatizados comercialmente disponibles, los cuales tienen implementado este formato. De esta forma, es posible beneficiarse de los últimos avances de la tecnología de computación.

Este formato posee fundamentalmente tres elementos: la estructura del registro, la designación del contenido y los datos del registro.

➤ **Estructura del registro**

La estructuración del registro permite controlar la representación de los diferentes datos almacenándolos en forma de una cadena de caracteres, conocidos como campos. Estos campos se identifican por tres caracteres numéricos llamados "tags".

La estructura del registro es una implementación de la norma ISO 2709-1981 (Format for bibliographic information interchange on magnetic tape) y la norma ANSI Z39.2 (American National Standard for Bibliographic Information Interchange). Dicha estructura utiliza etiquetas de registros y directorios.

➤ **Designación del contenido**

Se trata de los códigos y convenciones establecidos para identificar explícitamente los datos y caracterizarlos posteriormente dentro de un registro, y permitir la manipulación de los mismos.

Toda la información contenida en un registro con formato MARC es almacenada en forma de caracteres, codificados en formato ASCII extendido

Datos del registro

El contenido de los datos está definido por estándares externos al formato; como las reglas angloamericanas de catalogación, los encabezamientos de materia de la Biblioteca del Congreso Norteamericano, la clasificación de la Biblioteca Nacional de Medicina de EE.UU.

Un registro MARC está estructurado en campos fijos y campos variables.

Campos fijos

Está conformado por dos elementos:

◆ **EL LÍDER O GUÍA**

contiene valores codificados los cuales se identifican por la posición de cada carácter. Tiene una longitud fija de 24 caracteres y se encuentra al comienzo del registro. Estos datos definen los parámetros para procesar el registro.

el directorio

contiene el tag, la ubicación de comienzo del registro y la longitud de cada campo. La longitud del directorio es de 12 caracteres. El mismo finaliza con un carácter indicador de fin de campo.

◆ CAMPOS VARIABLES

El contenido de los datos está dividido en campos variables, que son datos alfanuméricos de longitud variable. Cada campo está identificado por tres caracteres, llamado "tag".

Según la norma ANSI Z39.2, el tag debe estar formado por caracteres gráficos ASCII alfabéticos o alfanuméricos. En el caso del formato MARC se utilizan los tags numéricos.

Los campos variables están agrupados en bloques según el primer carácter del tag, el cual identifica la función de los datos en el registro, como se muestra en la continuación.

- Bloque de formato bibliográfico

- 0XX = Número y código
- 1XX = Entrada principal
- 2XX = Títulos, edición, pie de imprenta
- 3XX = Descripción física
- 4XX = Mención de serie
- 5XX = Notas
- 6XX = Campos temáticos
- 7XX = Entradas secundarias, de apuntadores
- 8XX = Entradas secundarias de serie, existencias y localización
- 9XX = Reservado para implementación local

- Bloque de formato para existencias

- 0XX = Número y código
- 1XX = No definido
- 2XX = No definido
- 3XX = No definido
- 4XX = No definido
- 5XX = Notas
- 6XX = No definido
- 7XX = No definido
- 8XX = Existencias, localización y notas
- 9XX = Reservado para implementación local

Cada tag posee indicadores, los cuales contienen valores que interpretan o añaden información a los datos del campo. El formato MARC especifica dos posiciones de indicador al comienzo de cada campo variable, los cuales se definen independientemente.

Asimismo, existen los códigos de subcampos, los cuales identifican los elementos de datos en un campo que requiere o pueda requerir una manipulación por separado. Estos subcampos están conformados por dos caracteres; un delimitador y un elemento identificador.

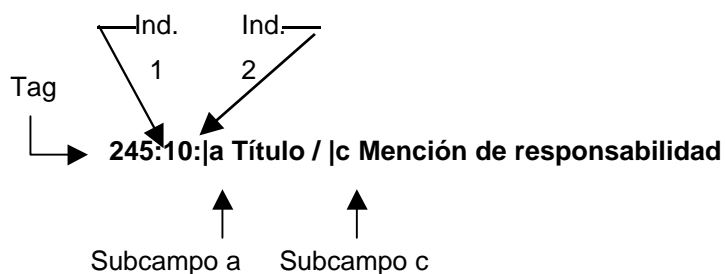


Fig. 1. Ejemplo de los elementos en los campos variables.

Como hemos podido observar, el contenido de los registros MARC está estrictamente formulado y, para usarlo uniformemente, los catalogadores requieren entrenamiento y experiencia. No obstante, MARC ha sido el más duradero y desarrollado de los formatos de metadada. No en vano Priscilla Caplan se refería a una forma más neutra de catalogar, cuando introdujimos el término metadada.

Como decíamos al comienzo, los registros bibliográficos, los cuales existen en el mundo de las ciencias de información, desde hace mucho tiempo son esencialmente metadada, con la diferencia de que hoy en día los estamos considerando en un contexto de recursos electrónicos. En vista de que estos recursos de información electrónicos se multiplican más rápido de lo que los indizadores y catalogadores profesionales puedan manejar con los métodos y sistemas existentes, muchos son los interesados en obtener una vía para mejorar la recuperación de información en la Web. Esta necesidad ha llevado a varios grupos de disitintos profesionales a reunirse con el objeto de mejorar esta situación. Como una muestra tenemos [4]:

- Los seis talleres de Dublin Core realizados hasta la fecha
- La conferencia de IEEE sobre avances en bibliotecas digitales (IEEE Advances in Digital Libraries Conference) celebrada en abril de 1998
- Las conferencias de IEEE sobre Metadata (IEEE Metadata Conferences)
- Los grupos de trabajo sobre biblioteca digital colaborativa (Digital Library Collaboratory Working Groups)
- Seminarios organizados regularmente por OCLC "Understanding and Using Metadata"

Instituto sobre "Managing Metadata for the Digital Library: Crosswalks or Chaos", copatrocinada por la Association for Library Collections and Technical Services (ALCTS) y Library and Information echnology Association (LITA), realizada en mayo de 1998.

El proyecto de metadata más conocido es el de *Dublin Core Metadata Set*. Lleva este nombre por cuanto el lugar de encuentro de bibliotecólogos y archivistas, investigadores, ingenieros de computación e información, editores, desarrolladores de softwares y miembros de los grupos de trabajo de Internet Engineering Task Force (IETF), fue Dublin, Ohio (EE.UU.) en marzo de 1995. El mismo fue patrocinado por Online Computer Library Center (OCLC) y el National Center for Supercomputing Applications (NCSA).

El objetivo era definir una hoja de instrucciones de forma que autores y editores de un documento pudiesen crear sus propios registros de metadata en un servidor de red, sin que ello representase un gran esfuerzo y dedicación. Este taller reconoció que la amplia indización y control bibliográfico de recursos Internet dependen de la existencia de un registro simple que describa los recursos en un ambiente de red, de manera que establecieron un grupo mínimo de elementos descriptivos.

Estos elementos corresponden a tres categorías, las cuales indican a grosso modo, la clase o alcance de la información allí almacenada: (1) elementos relacionados con el contenido del recurso; (2) elementos relacionados con el recurso al considerarlo como propiedad intelectual; (3) elementos relacionados con la descripción física del recurso [5]

A continuación presentamos una descripción de estos elementos:

Contenido del recurso	Propiedad intelectual	Descripción física o electrónica
Título	Autor o creador	Fecha
Temática	Editor	Tipo
Descripción	Coautor	Formato
Fuente	Derechos	Identificador
Idioma		
Relación		
Cobertura		

- **Título:** El nombre dado al documento electrónico
- **Temática:** El tópico tratado o asunto del documento. Se intenta promover el uso de un esquema de vocabulario controlado
- **Descripción:** Descripción textual del contenido del recurso, incluyendo el resumen en el caso de objetos del tipo documental o descripción del contenido en el caso de recursos visuales
- **Fuente:** El documento, ya sea impreso o electrónico, de donde se deriva el recurso, en caso de ser aplicable
- **Idioma:** El idioma del contenido intelectual del documento
- **Relación:** Relación con otros objetos, como por ejemplo imágenes en el documento, capítulos en un libro
- **Cobertura:** Ubicación espacial y duración temporal del recurso
- **Autor o creador:** Persona u organización responsable de la creación del contenido intelectual del recurso
- **Editor:** Entidad responsable de la disponibilidad del recurso
- **Coautor:** Persona u organización, además el autor principal, que han contribuido significativamente con la realización del recurso
- **Derechos:** Información sobre los derechos de autor o copyright
- **Fecha:** Una fecha asociada con la creación o disponibilidad del recurso
- **Tipo:** La categoría o género del recurso, como novela, poema, artículo, informe técnico, etc.
- **Formato:** La manifestación física del recurso electrónico, como HTML, Word, postscript, XML, xls, etc.

- Identificador: Serie o número usado para identificar el recurso, como URL, ISBN, etc.

Cada uno de estos elementos es opcional y repetible. Cada elemento puede igualmente contener subelementos para calificar o explicar el contenido del documento. Otra de las características de este estándar de metadata propuesto es que es sintácticamente independiente; la información puede ser representada en formato MARC, HTML, SGML, u otro.

Hasta abril de 1999 [6], el Dublin Core Element Set ha sido traducido a los siguientes idiomas: alemán, árabe, checo, chino, coreano, danés, español, finlandés, francés, griego, holandés, indonés bahaísmo (Bahasa Indonesia), italiano, japonés, noruego, portugués, thai y turco.

Si los datos anteriores los relacionamos con al formato MARC, los mismos se representarían de la siguiente manera:

Designación	Dublin Core	Formato MARC	Repetible
Título	"D.C:Title"	245 a	no
Temática	"D.C:Subject"		sí
Término no controlado		653 a	
Término encab. materia		650 a	
Clasif. Bib. del Congreso		050 a	
Clasif. Dewey		082 a	
Descripción	"D.C:Description"	520 a	no
Fuente	"D.C:Source"	786 t	
Idioma	"D.C:Language"		
Nota de idioma		546 a	
Idiomas		041 a	
Relación	"D.C:Relation"	787 n	si
Cobertura	"D.C:Coverage"		
Nota general		500 a	

Material cartográfico		255 c	
Tipo de informe		513 b	
Autor o creador	“D.C:Contributor”		no
Autor personal		100 a	
Autor corporativo		110 a	
Editor	“D.C:Publisher”	260 b	
Coautor	“D.C:Contributor”		sí
Coautor personal		700 a	
Coautor corporativo		710 a	
Otro coautor		720 a	
Derechos	“D.C:Rights”	506 a	
Fecha	“D.C:Date”	260 c	
Tipo	“D.C:Type”	516 a	
Formato	“D.C:Format”	538 a	
Identificador	“D.C:Identifier”		
Otro identificador estándar		024 a	
URL		856 u	
ISBN		020 a	
ISSN		022 a	

A manera de conclusión

Los profesionales de las ciencias de la información cuentan con los elementos para proponer ideas en cuanto a las mejoras necesarias para la recuperación de información en forma electrónica.

En el plano nacional, es necesario realizar discusiones sobre la base de lo discutido en eventos internacionales

Necesario interactuar conjuntamente con los profesionales de la informática de manera que haya agregación de valor en la suma de dos pericias, y no se reinvente la rueda por ninguna de las partes

Ahora bien, para que un esquema de metadata tenga éxito, debe contemplar un estándar tanto desde el punto de vista de su función descriptiva como su estructura y terminología, pues esto afectará el resultado de una búsqueda. Si todos los documentos contienen los mismos campos de búsqueda y en forma estandarizada, y utilizan el mismo vocabulario controlado, entonces el resultado de la búsqueda será más efectivo.

Z39.50

La necesidad de acceso global a la información está poniendo en contacto a bibliotecólogos e informáticos a fin de establecer tecnologías y estándares para el acceso indiscriminado a bases de datos de información documental y/o referencial. Parte importante de este proceso de globalización de la información se apoya en la definición de normas y estructuras comunes para el registro de los datos. En la primera parte de este artículo hicimos referencia al formato **Marc** y a los esfuerzos de definir un conjunto de campos a ser utilizados como metadata de los documentos, objetos y colecciones a los cuales se tiene acceso a través de la Web.

La proliferación de bases de datos en la Web obliga al establecimiento de procesos que faciliten a los usuarios finales el acceso a las mismas. En los actuales momentos cada base de datos está provista de un motor de búsqueda dotado con un lenguaje de recuperación al cual se accede a través de un interfaz de consulta cuya presentación y características propias dependen, en gran medida, de los parámetros de diseño de la institución oferente y a las facilidades proporcionadas por el sistema gestor de base de datos donde se almacena la información. Entonces, un usuario final debe interpretar y adaptarse a las modalidades que ofrece cada interfaz de consulta.

Cuando un usuario necesita realizar una búsqueda a través de la Web sobre las informaciones contenidas en distintas organizaciones, tiene que navegar a través de una gran cantidad de datos; entonces, recorre uno a uno los servidores que pudieran interesarle, con los problemas de diferencias de interfaz e inconsistencias ya mencionados.

Para homogeneizar esta diversidad surge el protocolo Z39.50, conjunto de reglas que constituyen un estándar para la localización de recursos de información en ambiente distribuido.

Este estándar provee los lineamientos para la comunicación Cliente-Servidor a efectos de la recuperación, a través de la red de computadoras Internert, de informaciones alojadas en diferentes servidores. Contiene los procedimientos y las estructuras requeridas para que un cliente pueda realizar búsquedas sobre las bases de datos ofrecidas por un servidor; recuperar los registros identificados por la búsqueda, recorrer listas de términos y ordenar los resultados de la consulta. También soporta control de acceso, control de recursos, servicios extendidos y facilidades de ayuda.

El protocolo dirige la comunicación entre aplicaciones de recuperación de información desde un cliente (origen) hacia un servidor (destino); no contempla la interacción entre el cliente y el usuario final.

La norma Z39.50 se propuso originalmente en 1984 para el acceso a información bibliográfica. A medida que el interés sobre el mismo se amplió, se estableció en 1990 el Grupo de Implementadores del Z39.50 (ZIG). Los miembros incluyen: fabricantes, vendedores, consultores, proveedores de información y universidades, quienes desean acceder o proveer acceso a varios tipos de información, incluyendo bibliográfica, texto, imagen, financiera, de utilidad pública, química y noticias. La asociación al ZIG está abierta a todas las partes interesadas.

El protocolo Z9.50 puede implementarse en cualquier plataforma por lo que permite que distintos sistemas informáticos puedan interoperar dando como resultado que el usuario final emplee un único interfaz para acceder a múltiples sistemas, con los comandos, formatos y estilos de presentación que le sean familiares.

Bases del Protocolo

El protocolo especifica los formatos y procedimientos que gobiernan el intercambio de mensajes entre un cliente y un servidor, permitiendo que una computadora cliente solicite a un servidor la búsqueda sobre una base de datos, la identificación de los registros que cumplen un criterio de específico y la recuperación de alguno o todos los registros identificados.

En esencia el la interacción bajo Z39.50 opera de la siguiente manera:

1. El cliente envía una solicitud a un servidor indicando una o más bases de datos e incluyendo una fórmula de búsqueda y algunos parámetros para determinar si los registros identificados por la búsqueda deberán ser enviados como parte de la respuesta.
2. El servidor responde indicando la cantidad de registros identificados y, posiblemente, con alguno o todos los registros.
3. El cliente puede entonces recuperar selectivamente los registros, asumiendo que éstos constituyen un “conjunto de resultados” y cada registro puede ser referenciado por su posición dentro del conjunto.

Las capacidades opcionales incluyen:

1. El cliente puede establecer las condiciones como desea recibir los registros. Por ejemplo, podría especificar: “si se recuperan hasta 5 registros transmita los registros en formato completo; de otra forma, transmita los registros en formato resumido”
2. El cliente puede indicar la sintaxis preferida con la que desea recuperar los registros, por ejemplo, USMARC
3. El cliente puede darle un nombre al resultado de la consulta para referencias posteriores
4. El cliente puede eliminar un conjunto de búsqueda ya nombrado
5. El servidor puede imponer restricciones de control de acceso, solicitando una autenticación antes de procesar la solicitud
6. El servidor puede proveer un control sobre los recursos, mediante el envío de un reporte de estado.
7. El servidor puede suspender el proceso y permitir al cliente la indicación de si desea continuar.

Modelo de una Base de Datos

La forma como se implementan las bases de datos a través de la Web varía considerablemente: diferentes sistemas adoptan diferentes estilos para describir el almacenamiento de los datos y sus métodos de acceso. Z39.50 establece el uso de un modelo abstracto común para describir las bases de datos, en el cual sistemas individuales pueden reflejar su modalidades de operación. Esto es, la

comunicación entre sistemas se realiza bajo términos estándar y comprensibles por ambas partes, con el propósito de búsqueda y recuperación de información desde una base de datos.

Al interior del estándar, cada registro es una colección de información relacionada, tratada como una unidad. El término *registro de una base de datos* se refiere a una estructura local de datos representando la información en un elemento en particular. Asociados a cada base de datos existen uno o más *puntos de acceso* que pueden especificarse para la búsqueda información. Un punto de acceso es una clave que puede ser utilizada bien sea sola o en combinación con otros puntos de acceso a través de operadores lógicos.

Búsquedas sobre bases de datos

Una fórmula de búsqueda se aplica a una base de datos, especificando los valores a ser localizados en los puntos de acceso (índices) a los datos. El subconjunto de registros obtenidos al aplicar la fórmula se denomina *conjunto de resultados*. Un conjunto de resultados puede ser referenciado en una fórmula de búsqueda subsiguiente y manipulado para formar un nuevo conjunto de resultados.

Una solicitud de búsqueda puede especificar una o más bases de datos e incluir una fórmula donde se mezclan puntos de acceso y operadores booleanos. Cada punto de acceso consiste de un término de búsqueda y unos atributos. Los atributos cualifican el término y están representados por un par de variables que definen el tipo de atributo y el valor de ese tipo (por ejemplo, el tipo puede ser "punto de acceso" y el valor "autor"; o el tipo puede ser "truncación" y el valor "izquierda").

Recuperación de registros desde bases de datos

A continuación del proceso de búsqueda, el servidor Z39.50 pone a disposición del cliente el conjunto de resultados objetidos, para peticiones subsecuentes. Al solicitar la recuperación de un registros de un conjunto de resultados, el origen puede suplir un identificador de **esquema de base de datos**, una **especificación de elemento** y un identificador de **sintaxis de registro**.

A efectos de la recuperación de los registros de un conjunto de resultados, cada base de datos tendrá asociado uno o varios esquemas. Un esquema representa un acuerdo común compartido entre el origen

y el destino de la información contenida en los registros de la base de datos, para permitir la selección subsecuente de porciones de esa información vía una especificación de elemento.

Un esquema define una *estructura de registro abstracta* la cual, al aplicarla sobre los registros de la base de datos, resulta en un *registro abstracto de la base de datos*, que constituye una representación abstracta de la información en el registro. Una especificación de elemento aplicada a un registro abstracto de la base de datos resulta en una nueva instancia del registro abstracto de la base de datos. La especificación de elemento selecciona parte del registro abstracto de base de datos y también puede especificar formas variantes para esos elementos.

El destino aplica una sintaxis de registro a un registro abstracto de base de datos, obteniendo una estructura exportable referida como *registro recuperado*.

Modelo de un conjunto de resultados

En general, un proceso de búsqueda no requiere necesariamente de acceso físico a los registros; un conjunto de resultados de búsqueda se asume entonces como la identificación (ejemplo, apuntadores) a los registros, en oposición al conjunto de registros seleccionado por la búsqueda. Un conjunto de resultados puede ser utilizado como un mecanismo de selección para la transferencia de registros entre sistemas.

Ventajas del protocolo Z39.50

Son muchas las ventajas que para usuarios finales puede representar la aplicación de la norma Z39.50 en bibliotecas y centros de información. Enumeramos algunas de ellas:

1. Relación entre bases de datos con estructuras y diseños diferentes
2. Catálogos colectivos virtuales
3. Homogeneidad en los procesos de localización de información
4. Acceso transparente a diversas fuentes de información

Bibliografía:

- [1] Improving Access to Information and Services of Australian Governments. Working Group on Government Information Navigation, July, 1997. <http://www.nla.gov.au/lis/esd4.html>
- [2] Milstead, Jessica; Feldman, Susan. Metadata: Cataloging by Any Other Name ... En *ONLINE*, Enero 1999. <http://www.onlineinc.com/onlinemag/metadata/>
- [3] Caplan, Priscilla. You Call It Corn, We Call It Syntax-Independent Metadata for Document-Like Objects. En *The Public-Access Computer Systems Review*, 6 (4). 1995.
<http://info.lib.uh.edu/pr/v6/n4/capl6n4.html>
- [4] Milstead, Jessica; Feldman, Susan. Metadata. Metadata Projects and Standards.
<http://www.onlineinc.com/onlinemag/metadata/>
- [5] Taylor, Chris. An Introduction to Metadata. <http://www.library.uq.edu.au/iad/ctmeta4.html>
- [6] Weible, Stuart. The State of the Dublin Core Metadata Initiative April 1999. En *D-Lib Magazine*, 5 (4): abril 1999. <http://www.dlib.org/dlib/april99/04weibel.html>
- [6] Library of Congress Z39.50 Maintenance Agency Site. <http://lcweb.loc.gov/z39.50/agency/>
- [7] BIBLO TECH Information Technology for Libraries: "Z35.50: Part 1- an overview", July 1998
<http://www.biblio-tech.com/html/z39.50.html>
- [8] BIBLO TECH Information Technology for Libraries: "Z35.50: Part 2 – Technical Details", July 1998
http://www.biblio-tech.com/html/z39.50_part_2.html