

Normas en la Biblioteca Digital

Por E. García Camarero

Presidente de *SABINI automatización de Bibliotecas*.

Índice

Introducción

- 1.- De la automatización de bibliotecas a la Biblioteca Digital
- 2.- Comunicaciones y redes de ordenadores: el modelo OSI e Internet.
- 3.- La norma Z39.50
 - 3.1. Historia
 - 3.2. Ámbito de la norma
 - 3.3. Descripción de la norma Z39.50.
 - 3.4.- Funciones de la Z39.50
 - 3.5. Aplicaciones v usos de la norma
- 4.- Otras normas
 - 4.1.- Introducción
 - 4.2.- Los metadatos.
- 5.- Conclusión
- 6.- Bibliografía.

Normas en la Biblioteca Digital

Por E. García Camarero
Presidente de *SABINI automatización de Bibliotecas*.

Introducción

Cuando D. Pedro López me invito a participar en este seminario de la Universidad Menéndez Pelayo, el tema concreto que me propuso fue el de la norma Z39.50, y creo que por dos razones. La primera y principal, por la gran importancia y difusión que está obteniendo esta norma para facilitar el acceso a las bases de datos bibliográficos y a los catálogos de las principales bibliotecas del mundo y por tanto para difundirla en nuestro medio. La segunda menor, y evidentemente sustituible, se refería a mi cierta experiencia en la misma, al haberla utilizado técnicamente para el desarrollo de un proyecto de la Comunidad Europea, el proyecto ARCA, para la implementaron de un sistema cliente/servidor de búsqueda y recuperación de información bibliográfica en catálogos remotos.

Aceptando gustoso la invitación y el honor que ella representaba, le propuse sin embargo ampliar ligeramente el tema de mi comunicación, para situar dicha norma en el contexto en el que ha aparecido y se está desarrollando, la biblioteca digital, para destacar uno de los fenómenos contemporáneos que mas van a afectar a la difusión de las ideas y a acelerar el desarrollo científico y tecnológico, y en el que la actividad bibliotecaria jugará cada vez un papel mas importante.

Así pues, aunque me ocuparé principalmente de describir la norma Z39.50, también trataré, por una parte, del modelo OSI dentro del cual debe de enclavarse, y por otra, de la explosión de información digital, tanto libraria como no libraria, que empieza a estar a disposición de los usuarios a través de las redes informáticas y de los esfuerzos que se están realizando para su organización y fácil búsqueda y recuperación.

1.- De la automatización de bibliotecas a la Biblioteca Digital

En la automatización de las bibliotecas convencionales, el problema fundamental fue la construcción de catálogos automáticos, y a partir de los cuales ofrecer una serie de servicios a los usuarios para facilitar la lectura de las obras almacenadas en sus depósitos.

Las piezas de información (libros, revistas, otros documentos) residían fuera de los ordenadores; en ellos se alojaban solo los catálogos y los programas que facilitaban su creación y actualización, así como los programas de acceso a los catálogos para la búsqueda de información que permitiese la localización de los fondos requeridos. También facilitaban algunos otros servicios bibliotecarios como, principalmente, la gestión de la adquisición o del préstamo.

Los cambios substanciales que se han dado con el gran desarrollo de la informática y de las redes de ordenadores, están propiciando la aparición de la Biblioteca Digital, que se apoya en dos hechos esenciales: primero, en que las piezas de información pueden residir dentro de los ordenadores formando repositorios de información, y, segundo, en que estos repositorios pueden estar distribuidos por todo el mundo interconectados entre sí y accesibles desde cualquier ordenador personal.

La paulatina aparición de la Biblioteca Digital ha sido posible gracias al abaratamiento de los soportes de información, y con ello el del almacenamiento de grandes cantidades de información, y al aumento de la velocidad de proceso que ha dado paso a la aparición de procedimientos cada vez más finos y elaborados para el tratamiento de la información (con el consiguiente avance en el software de aplicación, en el desarrollo de sistemas operativos y de otras muchas herramientas informáticas). Brevemente diremos que todas estas causas han permitido la aparición y el crecimiento hasta su madurez de la industria de edición electrónica, a la que la Biblioteca Digital responde de igual forma a como respondieron las bibliotecas tradicionales ante la aparición de la imprenta.

Por otra parte, también hemos presenciado el gran desarrollo de las comunicaciones y contemplado como en los últimos años se producía una verdadera explosión en la intercomunicación de ordenadores, así como una extraordinaria proliferación de fuentes de información de todo tipo accesibles gracias a esa intercomunicación. Internet es el más vivo resultado de esta actividad.

Para ordenar y regular las grandes disparidades con las que se ha producido ese desarrollo se ha requerido la elaboración de normas, no solo para asegurar la calidad de los productos, sino principalmente para poner en orden elementos tan heterogéneos, y asegurar la interconectividad y comunicación entre los distintos agentes.

Así, junto a normas de jure respaldadas por organismos nacionales o internacionales competentes, se han utilizado también normas de facto más o menos impuestas por empresas o grupos comerciales al difundir sus productos. Un ejemplo de este enfrentamiento en el caso de normas de comunicación informática lo encontramos en las normas OSI frente a los protocolos TCP-IP utilizados por Internet.

2.- Comunicaciones y redes de ordenadores: el modelo OSI e Internet.

Cada sistema de comunicaciones utiliza una red física, una plataforma tecnológica, asociada al mismo. La estructura de estas redes tiene una gran importancia; unas permiten la comunicación interpersonal (como es el caso de la telefonía o telegrafía) y otras (como la radio, y la televisión), por el contrario, reducen a la mayoría de las personas a meros receptores de información. Las redes correspondientes al primer tipo tienen por característica esencial el hecho de que cada punto de la red es, a la vez, transmisor y receptor. Por el contrario, en las otras redes solo se emiten los mensajes desde pocos puntos y se reciben desde una infinidad de ellos; sin que sea posible convertir el receptor en transmisor y, por tanto, no se puede establecer el diálogo y ni el coloquio entre las personas que utilizan este último tipo de redes. Además, en ninguno de los dos casos se puede realizar el almacenamiento y conservación de la información, y esta se pierde una vez comunicada.

Las nuevas tecnologías de la información han permitido construir redes en las que se integran ambas estructuras; *Internet* es la red de este tipo mas difundida en todo el mundo; está formada por gran numero de ordenadores intercomunicados, que facilita a millones de personas recibir datos de la colectividad, al tiempo que esas mismas u otro gran número de personas pueden introducir, con facilidad, información, para que se distribuya por toda la red; en ella es posible tanto el dialogo y el coloquio como acceder a información almacenada y conservada en los ordenadores de la propia red.

2.1. Sistemas cliente/servidor.

La *red de ordenadores*, se va formando y creciendo paulatinamente (aunque cada vez con mayor velocidad) mediante la conexión telefónica de ordenadores, (libremente realizada por el deseo de sus propietarios) a un ordenador que ya esté en la red. Los ordenadores que componen la red son de dos tipos, unos llamados *servidores* y otros *clientes*. Esta diferenciación no viene determinada por las características físicas de cada ordenador, sino por la función que desempeña dentro de la red, es decir, por un software que lo convierte en tal. Los ordenadores *cliente* son utilizados por los usuarios de la red para enviar o recibir mensajes, y para ello deben de estar conectados a un *servidor*. Los ordenadores *servidor* gestionan la transmisión de mensajes que circulan por la red y que han sido pedidos o enviados por un ordenador *cliente* y son depositarios de toda la información disponible en la red. La red esta formada por varios *servidores* conectados entre sí; a cada uno de estos *servidores* están conectados varios ordenadores *cliente*.

Un *servidor* tiene una doble función: la de gestionar la comunicación y la de almacenar y gestionar la información disponible en la red. En el primer caso un *servidor* podría compararse con una oficina de correos en donde se reciben mensajes desde sus *clientes* o desde otros *servidores*. Los mensajes que recibe de sus *clientes* los retransmitirá a otro *servidor* a través de una ruta que pasa por otros varios hasta alcanzar el que esté conectado al *cliente* de destino. Los mensajes que recibe de otro *servidor* estarán

destinados a un *cliente* directamente conectado con él, o a ser retransmitido a otro *servidor* buscando la ruta que lleve el mensaje hasta el *cliente* final.

En cuanto a la función de almacenamiento, puede considerarse al *servidor* como un archivo local en el que se recibe información de cualquier *cliente* para almacenarla, organizarla, y hacerla asequible para atender las demandas que provengan de cualquier *cliente* de la red.

La conexión entre los ordenadores se realiza mediante líneas telefónicas usando normalmente las ofrecidas por las empresas telefónicas. Cada ordenador de la red se identifica por una dirección (unas señas), es decir cada ordenador tiene una dirección que hay que colocar en los mensajes para que llegue a su destino.

2.2. Conmutación de paquetes: protocolos de comunicación.

Toda la información que circula o está disponible en la red es información digital. En forma digital se puede codificar no solo texto, sino también imagen y sonido, lográndose así transmitir por la red cualquier tipo de información como escritos, imágenes fijas y en movimiento, y voz y sonido. En cuanto a su extensión los mensajes pueden variar desde una simple carta hasta ficheros informáticos de gran tamaño, no solo datos y programas de ordenador, sino también cualquier tipo de literatura, música, cine, etc. ...

En el uso tradicional del teléfono la comunicación entre dos usuarios se realiza logrando establecer temporalmente una línea física (compuestas por varios tramos que enlazan las distintas centrales telefónicas) entre los aparatos telefónicos de ambos; cuando hablan están actuando en tiempo real; una vez terminada la comunicación se abandona la línea utilizada para que se pueda recomponer otra línea para nuevos usuarios. Esta técnica se llama *conmutación de líneas*. En informática la técnica de comunicación es diferente. La información no se transmite en tiempo real por una línea previamente establecida, sino que siempre se hace en tiempo diferido (tal vez fracciones de segundo), y además antes de transmitirse sufre un proceso de encapsulamiento, de empaquetamiento, para facilitar su transmisión. Esta técnica se llama de *conmutación de paquetes*; consiste básicamente en incluir los mensajes en uno o varios (si el mensaje es muy largo) *paquetes de información*; cada uno de ellos debe de llevar, además del mensaje propiamente dicho, la dirección del destinatario y del remitente, así como otra información adicional sobre el tipo de información que contiene y sobre la forma en que se ha construido.

Cuando las condiciones de la red de comunicación lo permite se realizan los envíos de cada paquete a la dirección de destino, cada paquete puede recorrer caminos distintos hasta llegar a su destino y llegar en cualquier orden. Una vez recibidos en el destino todos los paquetes correspondientes a un mensaje se ordenan y se reconstruye este para dejarlo a disposición del destinatario en la forma en que lo envió el remitente.

Las reglas con que se forman en origen los *paquetes* y la información que se agrega para la buena gestión de la comunicación, deben de ser comunes en todos los emisores y conocidas por los receptores a la hora de reconstruir los mensajes. Estas reglas se conocen con el nombre de "*protocolos de comunicación*". El protocolo mas importante usado por Internet se denomina TCP-IP. La parte IP (Internet Protocol) suministra un

espacio uniforme de direcciones y de rutas para hacer llegar cada paquete a su destino a través de diversas redes físicas, pero no garantiza la llegada de todos los paquetes de un mensaje, ni cuando, ni en que orden. Para asegurar que todo se realiza correctamente está la parte TCP (Transfer Control Protocol), que se encarga de empaquetar la información y de asegurar una correcta reconstrucción del mensaje en destino, reclamando si fuera necesario repetir algún envío.

2.3. Los sistemas abiertos OSI

Esta posibilidad de transmitir paquetes de información de manera eficiente han permitido la aparición y facilitado el crecimiento de las redes de ordenadores. El hecho de conectarse entre sí un cierto número de ordenadores tiene un efecto multiplicador muy elevado en cuanto al rendimiento de los mismos.

Para que las actuales redes lograsen superar los inconvenientes que antes se presentaban a la hora de conectar ordenadores heterogéneos, o para hacer colaborar aplicaciones informáticas distintas, se ha hecho un gran esfuerzo de normalización que empezó con las bases teóricas establecidas por la Organización Internacional de Normalización (ISO) al definir el "Modelo de referencia para la interconexión de Sistemas Abiertos" (norma ISO 7498), conocido por su denominación abreviada como "Modelo OSI"

Aunque estamos familiarizándonos con el uso de Internet y comprobando que la comunicación por este medio es cómoda y fiable, no sabemos que esto es así gracias al empleo de protocolos que se ocupan de realizar las diversas tareas necesarias para el éxito de la comunicación. Esos protocolos se definen dentro de la estructura general establecida por el "Modelo OSI", que describe las tareas que se deben realizar para establecer la comunicación entre sistemas abiertos. Estas tareas se agrupan en siete *capas* o niveles (*layers*). Cada capa está compuesta por dos programas informáticos situados cada uno de ellos en un ordenador distinto (los dos ordenadores que se quieren comunicar), programas que se comunican entre sí usando el protocolo específico definido en cada *capa*. Estas siete capas se agrupan a su vez en dos categorías: las *capas de datos* (1, 2, 3, 4) y las *capas de información* (5, 6, 7). Las primeras capas están destinadas a tratar los datos en bruto, es decir, a establecer como deben circular los bits o los paquetes de bits por la línea de comunicación entre los ordenadores. Las *capas de información* intervienen cuando la información adquiere un significado específico para una aplicación informática o para un ser humano.

Las distintas capas tienen las siguientes funciones:

- *Capa 1 Física* (Physical Layer): es responsable de la transmisión de los bits en los aspectos mecánicos y eléctricos de la transmisión.
- *Capa 2 Encadenamiento de datos* (Data link layer): es responsable de construir los "paquetes" de bits.
- *Capa 3 Red* (network layer): responsable de la circulación de la información por la red, así como la segmentación de los mensajes en origen y el reagrupamiento en la recepción.
- *Capa 4 Transporte* (transport layer): Permite la transmisión de datos de forma eficaz y segura.

- *Capa 5 Sesión* (session layer): Permite el intercambio organizado y sincronizado entre las aplicaciones. Las aplicaciones se definen como reglas de dialogo, y esta capa es responsable de hacer respetar estas reglas.
- *Capa 6 Presentación* (presentation layer): Asegura que el contenido de la información sea respetado durante la transmisión. En este sentido se dice que esta capa es responsable de la sintaxis de las informaciones transmitidas.
- *Capa 7 Aplicación* (application layer): sobre este nivel se apoya la norma Z39.50 que trataremos a continuación.

3.- La norma Z39.50

La norma Z39.50 es una norma norteamericana (NISO/ANSI) en la que se da un conjunto de reglas y procedimientos para regular el comportamiento de dos sistemas informáticos que se comunican entre sí con la finalidad de la búsqueda y recuperación de información en bases de datos principalmente bibliográficas. En una red de ordenadores es una *norma abierta* (OSI), que permite la intercomunicación entre sistemas heterogéneos, es decir, entre sistemas que operan con hardware y software diferentes (es una norma que se sitúa en la *capa de Aplicación* del modelo OSI). El Z39.50 sustituye como norma internacional con el nombre ISO 23959 a la antigua norma SR aprobada en 1991 por ISO.

3.1. Historia

Puede decirse que el origen de la norma Z39.50 está en los esfuerzos que se hicieron en los años 70 para normalizar la comunicación entre un pequeño número de Bases de datos bibliográficos (que por otro lado eran bastante homogéneas) alojadas en instituciones como la Biblioteca del Congreso, OCLC (Online Computer Library Center), y en la red RLIN (Research Libraries Information Network). La principal finalidad de esta intercomunicación era facilitar la catalogación compartida entre esas instituciones. En el proyecto LSP (Linked System Project) se comenzaron a definir los protocolos de comunicación y a trabajar también en su implementación. Al comienzo de los años 80 solo se realizaban tareas de implementación y los problemas de definición formal de los protocolos pasaron a la National Information Standards Organization (NISO).

Con anterioridad, ya en 1979, funcionaba un comité NISO creado con esa finalidad formado por un pequeño (y cerrado) grupo de expertos que trabajaban de forma solitaria hasta que sus resultados finales se sometían a votación. Aunque estos resultados no se aprobaron en la primera votación de 1984, su trabajo se aprobó en 1987 como norma, con la denominación "American National Standard Z39.50, Information Retrieval Service Definition and Protocol Specifications for Library Applications". Era la primera versión de la norma y fue publicada en 1988 (versión que es conocida como Z39.50-1988). Esta norma se basaba en el marco establecido por el "Modelo OSI". Su implementación fue muy escasa ya que solo se hizo dentro del proyecto que la impulsó (LSP), y en una aplicación que tuvo cierta difusión conocida como WAIS (Wide Area Information Server)

A finales de los 80 los objetivos de la Z39.50 comenzaban a cambiar. La comunidad interesada en el intercambio y acceso a información bibliográfica era mucho mayor que el pequeño número de bibliotecas que dieron origen a la norma. Los sistemas de automatización de bibliotecas estaban implantados en infinidad de bibliotecas, funcionando con sus OPACs correspondientes y era posible el acceso remoto desde los despachos de profesores y profesionales. Estas facilidades hacían crecer la demanda de información bibliográfica. Varias normas concurrían para atender esa demanda y para facilitar intercomunicación bibliográfica. En particular la ISO había estado trabajando en la definición de la norma conocida como SR (Search and Retrieval, definida cada una por separado y reconocida como norma internacional en 1991 con los números ISO

10162/10163). Esta norma era casi idéntica a la Z39.50, salvo que usa otras reglas para definir la sintaxis y para la codificación básica y omite algunas funciones. En estos años la NISO encargó a la Biblioteca del Congreso de Washington de reunir y coordinar los borradores que se estaban produciendo en una revisión de la Z39.50 en vista de armonizar esta norma con la que se estaba desarrollando por ISO. Para ello se disolvió el antiguo comité de la NISO y se constituyó un grupo no oficial y autoseleccionado que se conoce como ZIG (Z39.50 Implementors Group) que estaba formado inicialmente por representantes de 15 organizaciones. El grupo creció rápidamente. Se abrió por primera vez una lista pública en Internet (foro) para facilitar la discusión y la revisión de la norma y con ello el interés a participar en la definición de la norma a un sector cada vez más amplio. La actividad de ZIG tuvo su primer efecto en la aprobación de una versión de la norma la Z39.50-1992 (versión 2) que había sido elaborada con la participación de un gran número de personas que trabajaban, en su implementación, en entornos diferentes, y contenía como subconjunto a las normas ISO 10162/10163. Las principales dificultades de esta versión provenían de que al situarse explícitamente dentro del marco del "Modelo OSI", las redes creadas con los protocolos de comunicación derivados de este modelo se desarrollaban con dificultad. Para lograr que el Z39.50 pasase de la teoría a la práctica, era necesario desplazarlo a entornos que funcionasen con TCP/IP, es decir bajo Internet. En este sentido se lanzó entre 1992-1993 un programa para la prueba de la interoperabilidad de Z39.50 (Z39.50 Interoperability Testbed) con el objeto de estimular el desarrollo de gran número de implementaciones interoperables con la norma Z39.50 que funcionasen sobre Internet. El resultado fue la implementación de cierto número de clientes y de servidores que demostraron su funcionamiento en un marco comercial, como lo es el de las exposiciones de la ALA en las que concurren tanto fabricantes de productos informáticos para bibliotecas como los gestores de bibliotecas que buscan soluciones para sus bibliotecas. Estas experiencias incidían en el trabajo del ZIG, que era cada vez más amplio y en el que se atendían las demandas de todos los implementadores. El resultado de esta actividad condujo a la versión 3 (Z39.50-1995). En esta versión se han incluido importantes cambios y a ella nos referiremos en el resto de esta comunicación.

3.2 Ambito de la norma

Después de haber expuesto sucintamente el desarrollo histórico de la norma Z39.50, veamos esquemáticamente su ámbito de funcionamiento.

La norma regula la manera en que debe hacerse la comunicación y la transferencia de información entre dos sistemas con la finalidad de búsqueda y recuperación de información, principalmente bibliográfica. Prescribe que esta comunicación debe realizarse mediante lo que se conoce como una *sesión Z* entre un *cliente Z* (origin) y un *servidor Z* (target). El *cliente Z* es un programa que puede tener una interfaz de usuario independiente o estar embebido dentro del OPAC de un sistema de automatización de bibliotecas y usar el interfaz de este; este particular no lo determina la norma. El *servidor Z* es un programa que se adosa al sistema de gestión de la base de datos que utiliza el catálogo. Así como el cliente Z puede actuar independientemente o dentro de un OPAC, el servidor Z siempre tiene que trabajar conjuntamente con la base de datos que gestiona el catálogo.

Suponiendo la situación de un *cliente Z* incluido dentro de OPAC, podríamos definir esquemáticamente un proceso típico de búsqueda mediante la norma Z39.50 como sigue:

- Un usuario selecciona a través del menú de su OPAC la *biblioteca destino* (target) en la que esté instalado el *servidor Z* que actuará sobre la base de datos remota.
- El usuario introduce los términos de la búsqueda en su OPAC
- El software del OPAC envía los términos de la búsqueda y los detalles de la biblioteca destino al *cliente Z* que está integrado como parte del sistema de automatización de bibliotecas.
- El *cliente Z* traduce los términos de la búsqueda en una *frase Z* conforme a la norma Z39.50 y contacta con el servidor Z de la biblioteca de destino.
- Se produce después del contacto una negociación particular entre el cliente y el servidor para establecer las reglas de la *asociación Z* entre los dos sistemas.
- El *servidor Z* traduce la *frase Z* en una consulta a la base de datos de la biblioteca destino y recibe la respuesta correspondiente
- El *cliente Z* recibe los registros bibliográficos obtenidos como resultado de la consulta.
- Los registros son presentados en la interfaz de usuario del OPAC de la biblioteca origen.

Los clientes Z pueden enviar las peticiones de búsqueda a varias bibliotecas simultáneamente, con el consiguiente gran ahorro de tiempo en la búsqueda, cuando es necesario consultar a varias bibliotecas. Para formular las consultas se pueden utilizar los operadores booleanos (y, o, no), así como los operadores de comparación (<, >, =), los de proximidad y los de truncación. También pueden realizarse las búsquedas en todo un registro, en un campo, etc.

Normalmente el formato utilizado para el intercambio de registros es el formato MARC. Los servicios complementarios definidos por la norma Z39.50 permiten ordenar los documentos; hacer el barrido de los índices para facilitar la búsqueda; salvar los resultados de una consulta; salvar la formulación de una consulta para su uso posterior; definir una cadencia periódica de búsquedas; actualizar bases de datos remotas; crear la especificaciones para la exportación de datos.

Además la Z39.50 permite llevar el control del acceso a las bases de datos permitiendo solo a las consultas autorizadas; hacer las tareas propias de una contabilidad del uso de los recursos, en caso que así se establezca; suministrar información sobre las bases de datos a las que acceden los servidores con los que tiene establecido el servicio.

Las implicaciones que la norma Z39.50 esta teniendo, y tendrá todavía más en el futuro, es enorme. Y aunque su difusión no sea tan rápida como ha sido la del WEB, sus consecuencias son mas estructurales y de mas alto alcance. Algunos de los efectos que tendran sobre los servicios y operaciones bibliotecarias son los siguientes:

OPAC: la mayoría de estos sistemas se construirán de acuerdo con la norma Z39.50; lo que facilitara al usuario el buscar información en las principales bibliotecas del mundo usando el mismo método de acceso.

Catalogación: Permite, por una parte, un método simple y eficiente para descargar registros bibliográficos de los catálogos de unas bibliotecas en otras, y para realizar la catalogación cooperativa; por otra parte la facilidad de consulta simultanea a varias

bibliotecas repartidas por todo el mundo puede considerarse como la realización del viejo sueño del catálogo colectivo universal (en este caso los catálogos colectivos son virtuales en el sentido de que no deben acumularse todas los registros bibliográficos en único depósito).

Préstamo interbibliotecario: Uno de los beneficios inmediatos de los catálogos colectivos virtuales es la facilidad que representa para el préstamo interbibliotecario la inmediata localización del ítem requerido. También la Z39.50, mediante sus servicios especiales, ayuda al préstamo interbibliotecario con la posibilidad de gestionar los envíos y realizar las cuentas en los casos que así se necesitase.

Diseminación selectiva de información: realizada por el propio usuario utilizando los servicios, que ya mencionamos, permiten guardar las consultas y realizarlas en intervalos de tiempo predeterminado.

Las primeras empresas con *productos Z* fueron los grandes suministradores de sistemas automatizados de bibliotecas, que suministraban tanto *clientes Z* como *servidores Z* como parte de sus sistemas de manera que quien usaba uno de los grandes sistemas entraba a pertenecer al club selecto de quienes accedían a todos los registros bibliográficos pertenecientes a todas ellas.

En la actualidad están apareciendo implementaciones de *clientes Z* autónomos (es decir con su propia GUI y sin depender de ningún sistema de automatización de bibliotecas), de bajo costo y en algunos casos suministrados de forma gratuita, lo que ampliará de forma notable el número de los miembros del citado club, pero que no por numeroso perderá su calidad de selecto.

3.3. Descripción de la norma Z39.50.

La norma Z39.50 se describe en un documento oficial titulado "Information Retrieval: Application Service Definition and Protocol Specification", que seguiremos para exponer el modelo y las características generales del servicio de recuperación de información que en él se define con detalle.

Con la estructura específica que tienen los formatos y con los procedimientos mediante los que se regulan el intercambio de mensajes entre un cliente y un servidor, el cliente puede pedir al servidor que haga una búsqueda en una base de datos seleccionada por él, que identifique en ella los registros que responden a los términos indicados en la consulta y que recupere alguno o todos los registros identificados.

El objetivo de esta norma es facilitar la interconexión abierta de clientes y servidores para aplicaciones en las que los clientes buscan y recuperan información de las bases de datos de los servidores, que en general están situados en lugares remotos. Por tanto, teniendo en cuenta que las bases de datos son heterogéneas y que en su implementación difieren notablemente de unas a otras (los sistemas diferentes tienen diferentes estilos para describir el almacenamiento de los datos y los medios por los cuales se puede acceder a ellos) se necesitaba un modelo común y abstracto, para describir las bases de datos, al que pudiesen referirse los que proyectasen la implementación de un sistema particular. Con esto se

consigue que diferentes sistemas puedan comunicarse mediante términos comprensibles mutuamente gracias a la norma. También en la norma se describen los modelos de búsqueda y recuperación en una base de datos.

Búsqueda en una Base de datos.

En una búsqueda se especifican la o las bases de datos a consultar y los datos necesarios para realizar una consulta. El tipo de consultas que se usan para la búsqueda de información bibliográfica es la que en esta norma se define como de *tipo 1*, y consisten en expresiones que contienen uno o varios puntos de acceso vinculados por operadores lógicos. Por ejemplo:

"En la base de datos llamada "Libros" encontrar todos los registros para los que en el punto de acceso "palabra del título" aparece el valor "evangeline" Y en los que en el punto de acceso "autor" aparezca el valor "longfellow".

Cada punto de acceso se define como una secuencia de términos de búsqueda, con uno o varios atributos. Los atributos califican al término; cada atributo consiste en un par de elementos que representan un *tipo de atributo* y un *valor* de ese tipo (por ejemplo, tipo puede ser "punto de acceso" y valor "autor"; o tipo "truncación" y valor "izquierda".)

El subconjunto de los registros obtenidos como resultado de aplicar la pregunta a la base de datos se llama *conjunto resultado* (result-set). Una consulta posterior puede hacer referencia a un *conjunto resultado* y utilizarlo para formar un nuevo conjunto resultado.

Recuperación de registros de una base de datos.

Después de realizado el proceso de búsqueda, el *servidor Z* (destino/target) pone a disposición del *cliente Z* (origen) la información contenida en el *conjunto resultado* mediante peticiones adicionales de recuperación. Cuando hace una petición de recuperación de registros del conjunto resultado, el origen debe suministrar al servidor un identificador de esquema de base de datos, una especificación de elemento y un identificador de la sintaxis del registro.

Con objeto de recuperar registros a partir de un *conjunto resultado*, cada base de datos tiene asociado uno o mas *esquemas*. Un *esquema* representa una forma común de entender la información contenida en los registros de la base de datos compartida por el *cliente* y el *servidor* para permitir la selección de segmentos de información del registro por medio de una especificación de elemento.

Un *esquema* define una estructura abstracta de registro que cuando se aplica un registro de una base de datos se obtiene un registro de la base de datos abstracta que es una representación abstracta de la información contenida en el registro. Una *especificación de elemento* aplicada a un registro de base de datos abstractas se convierte en una instancia del registro de la base de datos abstracta. La *especificación de elemento* selecciona los elementos desde el registro de la base de datos abstracta, y puede también especificar algunas formas variantes de aquellos elementos.

El *servidor* (target) aplica una *sintaxis de registro* a un registro de base de datos abstracta convirtiéndose en una estructura exportable conocida como la de un registro recuperado.

Modelo del conjunto resultado

Se supone, en general, que un proceso de consulta no requiere necesariamente acceder a los registros físicos; de manera que un conjunto resultado contiene la identificación (por ejemplo mediante los punteros) de los registros, en vez del conjunto de registros físicos, seleccionados por la consulta. Un conjunto resultado puede usarse como un mecanismo de selección en la transferencia de registros entre sistemas; pero el propio *conjunto resultado* debe considerarse como una estructura local de datos y no es transferido (esto es, los registros son transferidos, pero no los punteros locales a los registros).

Con relación a la recuperación de registros, la estructura lógica de un *conjunto resultado* está formada por una lista ordenada de ítems, a la que para su reconocimiento y posterior uso se le asigna un nombre. Cada ítem de la lista está compuesto por una terna de elementos consistentes en :

- (a) Un número ordinal que corresponde a la posición de la terna en la lista.
- (b) Un nombre de base de datos (recuérdese que desde un servidor se puede acceder a varias bases de datos)
- (c) Un identificador único (solo con significado local) de un registro de la base de datos cuyo nombre se da en el segundo elemento de la terna.

Un ítem dentro del conjunto resultado se identifica unívocamente por la posición de su terna en la lista, es decir por su primer componente.

Cuando, con finalidad de búsqueda, se usa un *conjunto resultado* como un operando en una consulta, la estructura lógica es una de las siguientes:

- Modelo básico Un conjunto de pares, cada uno consistente en los elementos (b) y (c) del anterior modelo de recuperación.
- Modelo ampliado Un conjunto de ternas, cada una consistente en los elementos (b) y (c) del anterior modelo de recuperación, mas la inclusión de información no especificada asociada con cada registro, que puede ser usada información adicional en la búsqueda

Modelo de los Servicios Especiales

Entre los servicios Z39.50 se incluyen los Servicios Especiales (Extended Services (ES)) que son una clase de servicios reconocidos por esta norma aunque no son servicios Z39.50 propiamente dichos. El servicio ES es un servicio Z39.50, y una operación ES se produce como resultado de la inicialización de una tarea complementaria, que no es considerada parte de la operación ES Z39.50.

Una operación ES se inicia por el cliente (origen) mediante una petición ES. La respuesta de ES que completa la operación no es necesariamente una señal de que la tarea se haya terminado, puede indicar por ejemplo que ha comenzado o que está en la cola (se puede especificar en la petición ES que la tarea se acabe antes de dar la respuesta ES). Una tarea ES puede tener una duración mayor que la *asociación Z*.

Ya vimos algunos ejemplos de servicios ES al citar las posibilidades de salvar un conjunto resultado, y exportar u ordenar un documento.

La norma también regula como se deben especificar las tareas ES, es decir indica como dada una de estas tareas quedará representada por un registro, llamado paquete de tareas, perteneciente a una base de datos especial (la "base de datos de servicios especiales") mantenida por el servidor (target). El origen mediante la petición de un servicio especial (ES) crea un paquete de tareas sobre la base de datos de servicios especiales. Esta base de datos puede ser consultada y sus registros recuperados mediante las facilidades de Búsqueda y Recuperación de Z39.50. El *cliente Z* (origen) puede buscar paquetes de tareas por varios criterios, como por ejemplo que sean de un determinado tipo, o hayan sido creados por un usuario particular, o por estén en un status particular (por ejemplo: pendiente, activa, o completo) entre otros.

Explain

El cliente (origen) puede obtener detalles de la implementación del servidor (target), como las bases de datos que incluye, o los conjuntos de atributos, conjunto de diagnósticos, sintaxis de los registros, y especificaciones de elementos que soporta. El origen obtiene estos detalles a través de la facilidad Explain Z39.50. El servidor mantiene esta información en una base de datos a la que el origen puede acceder mediante las facilidades de Búsqueda y Presentación de Z39.50.

Esta base de datos "explain" aparece al origen como cualquier otra base de datos soportada por el servidor, pero tiene un nombre bien conocido y una sintaxis de registro predefinida. También ciertos términos, que corresponden a categorías de información, están predefinidas para permitir un nivel semántico de interoperatividad. Cada categoría de información tiene su propia etiqueta y todas están incluidas en la sintaxis de Explain.

3.4.- Funciones de la Z39.50.

Después de describir los modelos y el marco de funcionamiento de la norma Z39.50, veamos algunas de las funciones que pueden realizarse desde esta norma, y cual es su estructura.

La norma describe 11 funciones, de las cuales las cuatro primeras pueden considerarse como el núcleo de ellas (las encargadas de realizar la conexión, la búsqueda, y la recuperación y cancelación de registros), las siguientes son complementarias y aumentan la elaboración de los datos obtenidos.

Las funciones son las siguientes:

Inicialización: mediante la que se regula la forma de establecer una *asociación Z* y se negocian los niveles del servicio.

Búsqueda: mediante la que se regula como se realiza el envío una consulta a la base de datos y se obtiene como repuesta un *conjunto resultado* y los primeros registros.

Recuperación: mediante la que se regula como se recuperan los registros del *conjunto resultado* especificados por el *cliente Z*.

Cancelación de un conjunto resultado: mediante la que regula como se inicia la cancelación de *conjuntos resultado* almacenados en el servidor *Z*.

Barrido: mediante la que regula como se inicia, desde un cliente, un barrido en un índice de un servidor desde un determinado punto de acceso.

Clasificación: para regular la forma clasificar un conjunto resultado en un orden determinado.

Control de acceso: para regular el control de palabras de acceso y otros controles

Control de recursos y contabilidad: establece la forma de llevar la contabilidad, el control del credito, etc.

Explicar(Explein): mediante la que se regula la forma de acceder a una base especial del servidor *Z* en la que se deposita información sobre los detalles de implementación del servidor y sobre las bases de datos por el gestionadas

Servicios especiales: mediante los que se establecen la forma de iniciar "paquetes de tareas" externas como por ejemplo acceder a sistemas de Préstamo interbibliotecario.

Terminación: establece la forma de cerrar una *asociación Z*.

Veamos con un poco mas de detalle las características de las cuatro primeras funciones que forman el núcleo de la norma Z39.50.

Inicialización: El cliente *Z* contacta con un servidor *Z* y sugiere algunos parámetros básicos para la sesión. El servidor *Z* puede modificar estos parámetros y, si se ponen de acuerdo, iniciar una *asociación Z*. Entre la información que se intercambian para establecer una asociación están: la versión de Z39.50 utilizada por cada uno, opciones permitidas, numero de registros, otros servicios.

Búsqueda: Esta función es la mas importante de toda la norma y en la que aparece toda su potencia. Los parámetros empleados para describir los registros bibliográficos son los incluidos en el conjunto denominado *Bib-1*, podemos verlos en la tabla siguiente

Atributos de uso

Nombre personal

Nombre corporativo

Nombre de conferencia

Título

Título de Serie

Título Uniforme

ISBN

ISSN

Número de la LoC

Número de la BNB

Número BGF

Número local

Clasificación Dewey

Clasificación UDC

Clasificación "Bliss"

Clasificación local

Encabezamiento de Materia

Materias "Rameau"

Indice de materias BDI

| | |
|-----------------------------------|----------------------------------|
| Materias INSPEC | Autor-título |
| Materias MESH | Tipo de registro |
| Materias PA | Nombre |
| Encabezamiento Materia LoC | Autor |
| Encabezamiento Materia RVM | Autor-nombre personal |
| Encabezamiento de Materia local | Autor nombre corporativo |
| Fecha | Autor nombre de Conferencia |
| Fecha de publicación | Identificados normalizado |
| Fecha de adquisición | Materias LoC (imfantil) |
| Título Uniforme | Nombre personal como materia |
| Título Colectivo | Cuerpo del texto |
| Título paralelo | Fecha/hora ingreso en db |
| Título de cubierta | Fecha/hora ultima modificacion |
| Resto del título | Autoridad/formato |
| Subtítulo | Concepto-texto |
| Funcionamiento del título | Concepto-referencia |
| Título (otra variante) | Cualquiera |
| Título anterior | Opciones de Servidor |
| Título abrevió | Editor |
| Título extendido | Registro-fuente |
| Subdivisión de materia | Editor intelectual |
| No. Bibliog. nacional. | Nivel bibliográfico. |
| No. depósito legal | Clase Geográfica |
| No. publicación oficial. | Indizado por |
| No. editor musical | Escala cartográfica |
| Numere db | Clave Musical |
| Código de idioma | Periodo referido |
| Código de área geográfica | Numero de Informe |
| Código de institución | Número de stock |
| Nombre y título | Número Temático |
| Nombre geográfico | Tipo de Material |
| Lugar de publicación | Identificador de Documento |
| CODEN | Tipo de contenido |
| Generación de microforma | En cualquier parte |
| Resumen | Autor-título-materia |
| Nota | |
| Atributos de la relación | |
| menor | regExpr-1 |
| menor o igual | regExpr-2 |
| igual | |
| mayor o igual | Atributos de integridad |
| mayor | subcampo incompleto |
| no igual | subcampo completo |
| fonético | campo completo |
| raíz | |
| relevancia | Atributos de posición |
| idéntico | comienzo del campo |
| | comienzo del subcampo |
| | cualquier posición en campo |
| Atributos del truncamiento | |
| truncamiento a la derecha | |
| truncamiento a la izquierdo | |
| izquierda y derecha | Atributo de la estructura |
| no truncar | frase |
| proceso # en búsqueda término | palabra |
| | clave |

| | |
|-------------------------|-----------------|
| año | estructura |
| fecha (normalizada) | texto libre |
| palabra de lista | número local |
| fecha (sin normalizar) | cadena |
| nombre (normalizado) | cadena numérica |
| nombre (sin normalizar) | |

La función de búsqueda se descompone de dos tareas o servicios: la *consulta* y la *respuesta*. El *cliente* emite un mensaje con la consulta. La anterior lista de atributos nos da una idea de la riqueza con que puede realizarse una consulta. Cuando el servidor procesa una consulta crea un *conjunto resultado* (con los punteros de los registros de la base de datos que satisfacen la consulta), y envía como respuesta al *cliente* un mensaje con información sobre los resultados de la búsqueda y unos pocos registros, de acuerdo con lo solicitado en la consulta.

Recuperación: ya vimos que mediante esta función se regula la forma en que se recuperan los registros del *conjunto resultado* especificados por el *cliente Z*. Esta función está compuesta de tres tareas o servicios: *presentación*, *respuesta* y *segmentación*. La *presentación* es una petición realizada al servidor para que envíe ciertos registros obtenidos en la función anterior mediante un mensaje en el que se pueden indicar: el número de registros pedidos (10 registros), la posición inicial a partir de la cual se piden los registros (a partir del número 25), intervalos adicionales (25-27, 30-32), indicar un nombre de conjunto resultado que hubiese sido salvado con anterioridad, el formato de los registros (por ejemplo IBERMARC), y si los registros pedidos deben ser segmentados se indica el tamaño de los segmentos. En la *respuesta* del servidor se transmiten al cliente los registros solicitados y la información adicional que fuera necesaria. La *segmentación* se refiere a la manera de fraccionar la información para una eficiente transmisión por la red, es pues una tarea en la que no se pide información especial al bibliotecario.

Cancelación de un conjunto resultado: Con esta función realiza la cancelación de *conjuntos resultado* almacenados en el servidor Z, tanto el actual creado en la última búsqueda, como alguno de los salvados en búsquedas anteriores. El cliente recibe una lista de los conjuntos salvados para indicar cuales quiere cancelar.

3.5. Aplicaciones y usos de la norma

Como ya hemos reiterado varias veces, la norma se aplica en la recuperación de registros bibliográficos contenidos en catálogos remotos mediante los programas que implementan a los llamados *cliente Z* y *servidor Z* descritos en la norma. El servidor debe de estar contenido en el sistema de automatización de bibliotecas, que en virtud de tener un *servidor Z*, se convierte en un *sistema abierto*; por tanto los servidores suelen ser desarrollados por los propietarios de los sistemas de automatización de bibliotecas que los contienen. Sin embargo, los clientes no tienen necesariamente que estar incluidos como

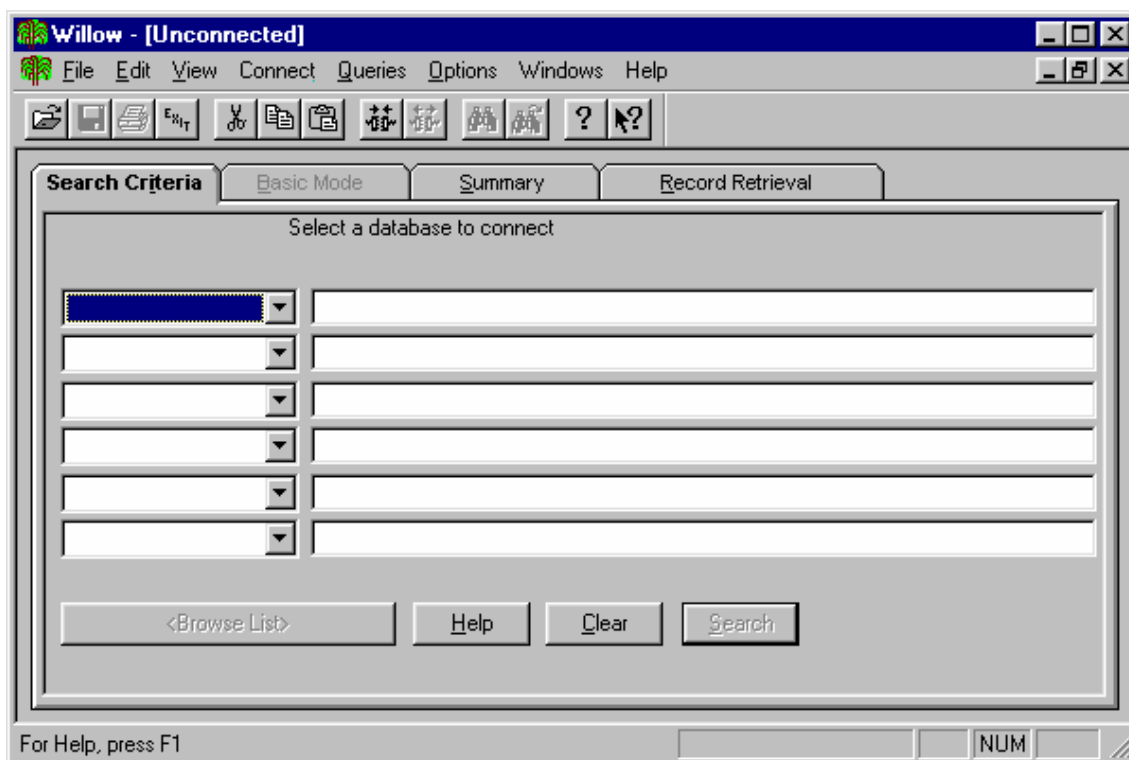
parte de uno de esos sistemas generales, y por tanto funcionar autónomamente y ser desarrollados independientemente de ninguno de tales sistemas.

Existen múltiples implementaciones de clientes Z39.50. Algunos de ellos se distribuyen por canales comerciales, desarrollados las empresas de los grandes sistemas automatización de bibliotecas como es el caso *GeoPac* (por GEAC Computers Inc., USA), *VTLS* (por VTLS Inc., USA), *WinPAC* (por Ameritech Library Services, USA), o por otras de empresas de documentación como *BookWere?* (Sea Change Corporation), *DRAFind* (Data Research Associates), o *Znavigator* (producido por Enware S.A. Madrid), otros se distribuyen de forma gratuita (feeware), como ocurre con CanSerach (por la Biblioteca Nacional de Canadá), Willow (por la Universidad de Washington), o de ICONE (desarrollado por la Comunidad Europea pero que se distribuye por la Biblioteca del Congreso de Washington).

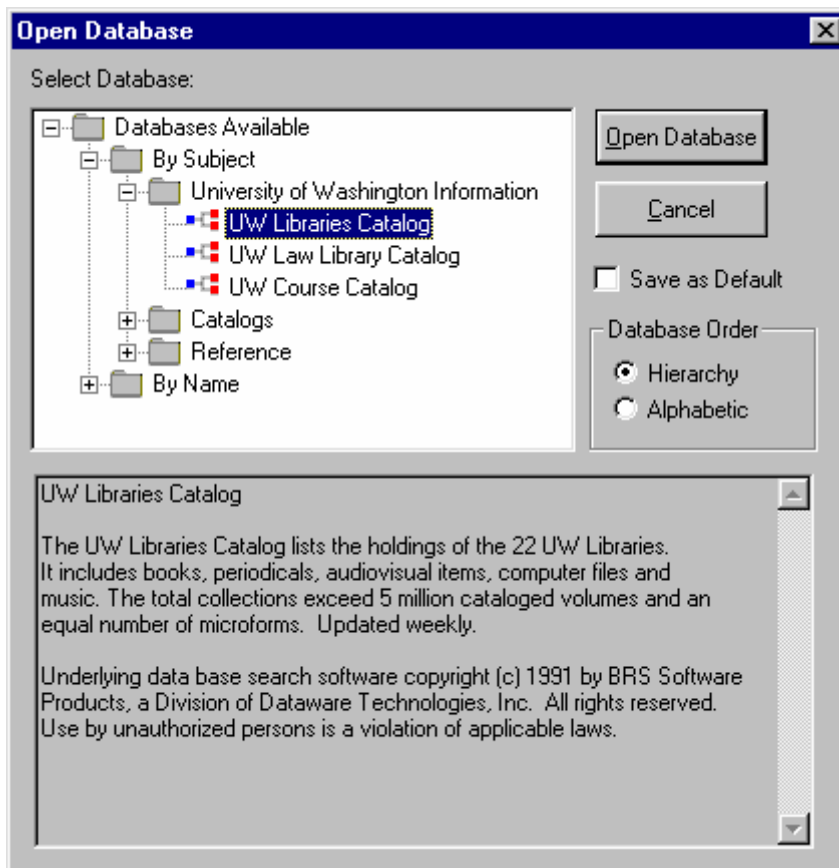
Por otra parte aunque los servidores los tienen que desarrollar los propietarios de los sistemas de automatización de bibliotecas, existen herramientas (kit de desarrollo) que facilitan esta tarea. En particular, mencionaremos que la Biblioteca Nacional de Canadá pone de acceso libre los programas fuente de su servidor. Citaremos también la existencia de un kit de desarrollo para servidores con la norma Z39.50, que puede obtenerse de forma gratuita : *Isite (Claringhouse for Networked Information Discovery and Retrieval)* y la dirección electrónica es <http://vinca.cniidr.org/software/Isite/>.

Veamos a continuación a manera de ejemplo, como son las pantallas de consulta de uno de los clientes Z, que hemos mencionado mas arriba.

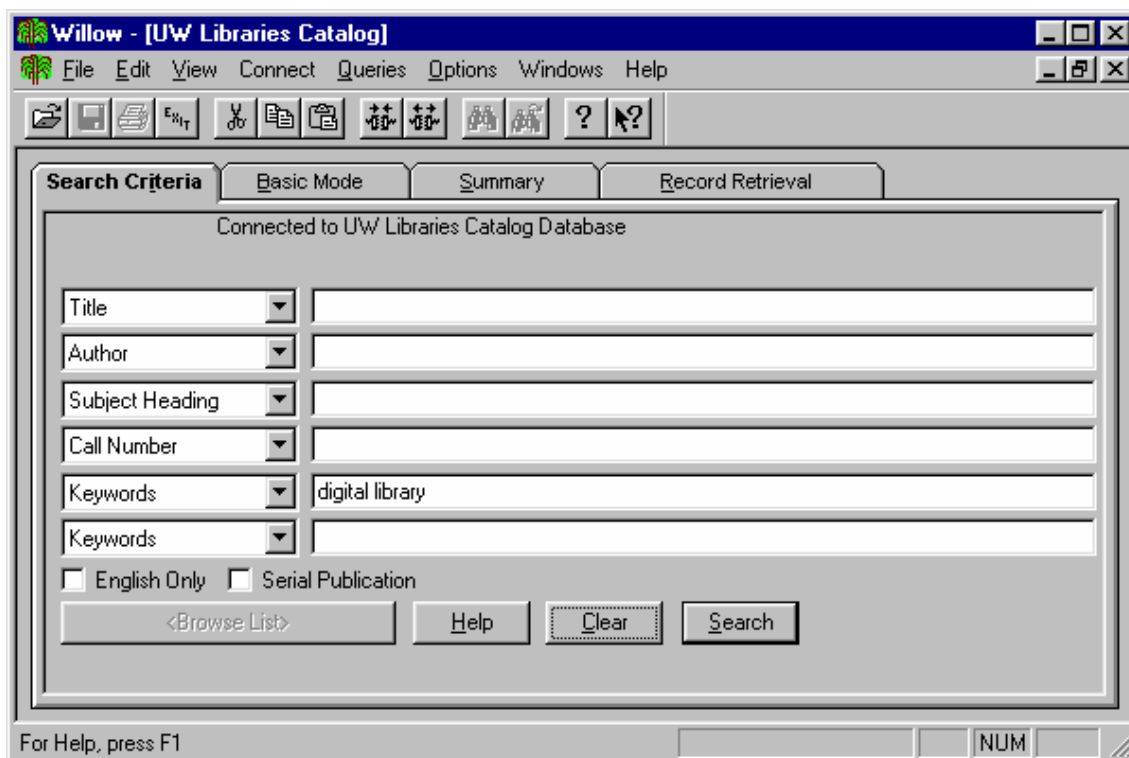
La pantalla inicial del cliente Willow es como sigue:



En la que aparecen implementados varios de las funciones que están descritas en la norma. La primera de ellas es la inicialización y la selección de una bases de datos, que se hace con ayuda de la siguiente ventana en la que se muestran, en este caso de forma jerárquica, las bases de datos accesibles.

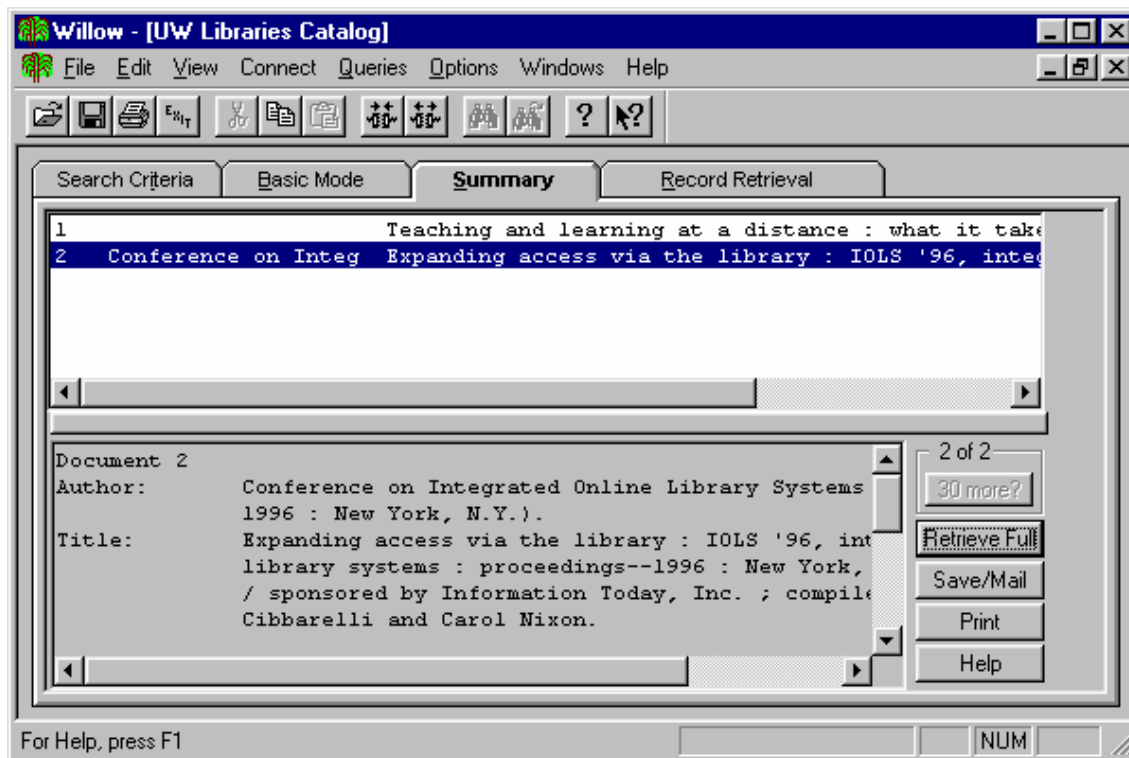


Después de seleccionada una base de datos, se retorna a la pantalla anterior para preparar los datos y realizar una consulta (Search). Observamos que las solapas nos indican distintas posibilidades. El modo básico (basic mode) nos facilita la consulta utilizando un barrido (por títulos, autor, materia, palabra clave), o mediante criterios de búsqueda (search criteria) o para mostrar con formato sucinto los registros encontrados en la base de datos, o bien recuperar el registro completo. En la pantalla que sigue mostramos una búsqueda con el criterio de palabra clave (keywords) para el valor "digital library", como vemos a continuación:

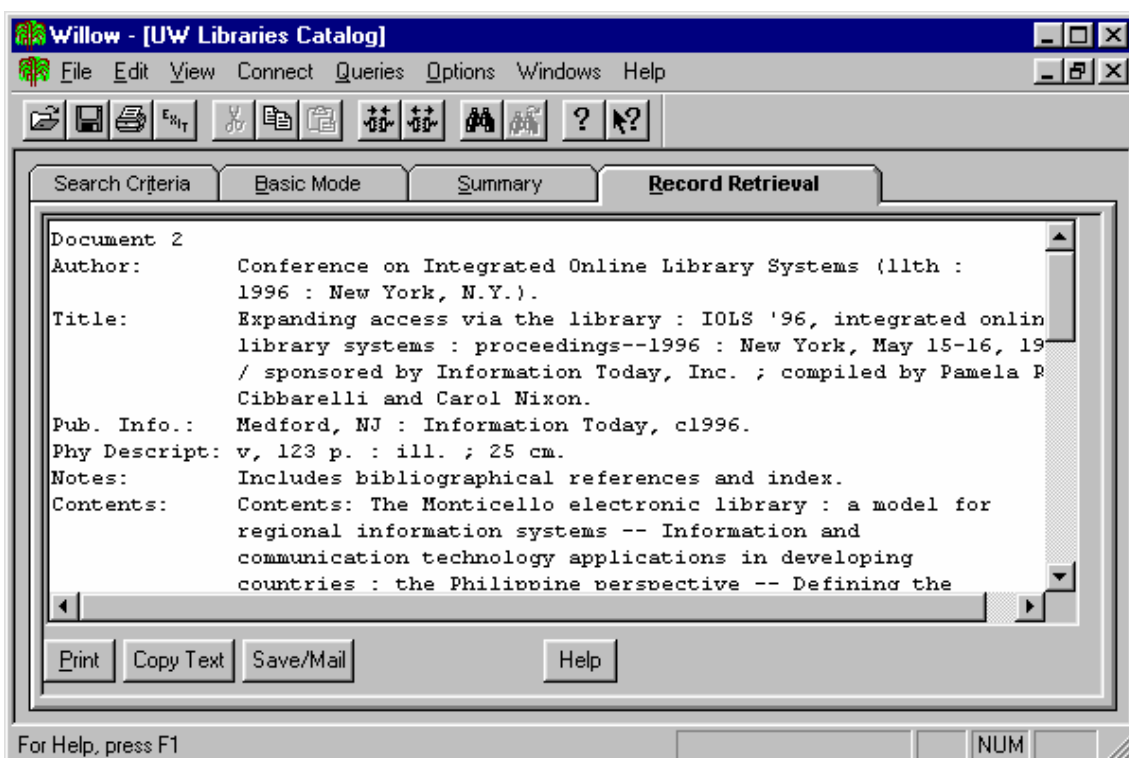


Observamos que se puede restringir a obras solo en inglés, o indicar que se trata de una publicación en serie. Al presionar el botón *Search* se realiza una búsqueda, cuyo resultado puede verse presionando la solapa *Summary*, como vemos a continuación:

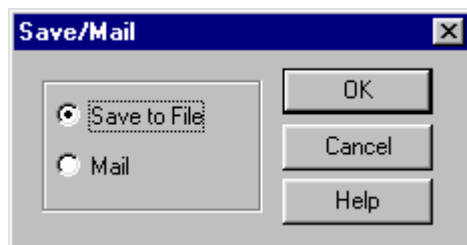
En la búsqueda hemos obtenidos dos referencias, a cada una de las cuales se dedica una línea, en la parte superior, para que sirvan de puntero a la hora de recuperar el registro completo. En la parte inferior aparecen las referencias bibliográficas en forma sucinta.



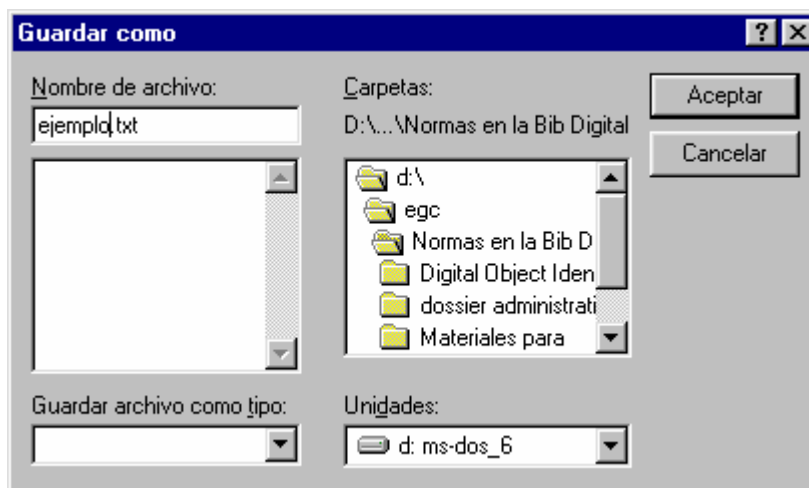
También en la parte inferior derecha aparece una columna de botones, el primero aparece debajo del numero registros recuperados y ofrece la posibilidad de obtener 30 más si los hubiere, el segundo se utiliza para recuperar los registros completos, el tercero para salvarlo en un fichero propio o para enviarlo por correo electrónico a la dirección que le indiquemos, el cuarto sirve para imprimir los resultados. Si usamos el botón *Retrieve full* obtenemos la siguiente pantalla en la que aparece el registro completo de la referencia seleccionada en la pantalla anterior:



En su parte inferior aparecen tres botones para imprimir, copiar el texto para poderlo pegar en otra parte, y salvarlo en un fichero propio o enviarlo por correo electrónico. Siguiendo este ejemplo pulsemos el botón *Save/Mail* y aparecerá la siguiente ventana auxiliar



Donde seleccionaremos la alternativa *Save to File* para salvarlo en un registro propio y accederemos a nuestro explorador Windows procediendo de la forma habitual



4. - Otras normas

4.1.- Introducción

Como hemos podido percibir en la sucinta descripción, hecha más arriba, de la norma Z39.50, esta se ha desarrollado para atender básicamente la necesidad de búsqueda y recuperación remota de información bibliográfica, entendiendo como tal a algo todavía muy impregnado de la edición impresa de monografías y publicaciones seriadas. Sin embargo, los nuevos desarrollos tecnológicos nos están conduciendo a una situación en la que gran parte de los fondos de las bibliotecas estarán formados por materiales en forma digital; y la estructura de estos materiales se alejará, en muchos casos, de la correspondiente de los actuales materiales librarios.

Así, los fondos de esa gran biblioteca virtual que se está formando, serán en unos casos colecciones de materiales en forma digital, bien convertidos a partir de materiales textuales impresos (libros, revistas, periódicos, etc.), de materiales gráficos (grabados, fotografías, etc.), de registros sonoros (obras musicales, discursos, programas de radio, etc.), o de imágenes en movimiento (cine, vídeo, televisión, etc.); o bien creados directamente en formato digital, como en el caso de las bases de datos (numéricos, estadísticos, cartográficos, etc.), o las muestras biológicas digitalizadas o provenientes de un scanner, las imágenes de satélite, los programas informáticos, los objetos de realidad virtual, y otra gran cantidad de diversos tipos de información capturada en la mayor parte de los casos de forma automática.

Los sistemas informáticos para gestionar la Biblioteca Digital, no serán como los actuales sistemas automáticos de bibliotecas que nos remiten a documentos impresos sobre papel, almacenados en las baldas de las estanterías de los depósitos de las bibliotecas tradicionales. Los nuevos sistemas informáticos, además de gestionar los catálogos o las bases de información, deberán poder almacenar los distintos tipos de documentos digitales (provenientes de diversas fuentes y en diversos formatos), y localizar y distribuir los documentos solicitados desde cualquier parte de el mundo.

Muchas de las características funcionales de las bibliotecas clásicas se mantendrán también en las bibliotecas digitales. Como son las de evaluar y seleccionar materiales digitales para su adquisición, las de conservar los materiales electrónicos de forma segura e inalterable (gestión de los depósitos); o las relativas a la descripción e indización de los nuevos materiales digitales incorporados (catalogación y clasificación) y su distribución entre los usuarios, y otras tareas administrativas en relación con los lectores (como el registro de autorización de uso, la facturación de los servicios, la diseminación selectiva de información etc.). Funciones todas ellas similares a las realizadas por una biblioteca clásica, aunque, eso sí, realizadas con otros procedimientos y técnicas cada vez mas automáticos que requieren, para su manejo, de un software muy sofisticado, de un hardware avanzado (formado por servidores, terminales, dispositivos especiales, mecanismos de comunicación, redes), y todo esto necesitará lo que, como siempre, es mas

importante: de un personal cada vez más capacitado para la integración eficiente de todos los recursos.

En la actualidad la enorme difusión del uso de los ordenadores personales de gran potencia y las nuevas aplicaciones propiciadas por su interoperabilidad a través de Internet, junto con el crecimiento y la disponibilidad de materiales digitales en gran número de servidores, ha abierto la posibilidad de considerar a toda esa información disponible (como los libros, revistas, documentos, imágenes o publicaciones depositados en cualquier servidor situado en cualquier parte del mundo), como el fondo de la Biblioteca Digital (considerada esta en su sentido más amplio).

Se considera a la Biblioteca Digital como una biblioteca sin paredes y sin salas de lectura, en la que sus fondos están disponibles en forma digital y son accesibles de forma remota desde los ordenadores de cualquier institución, o desde los ordenadores personales de uso doméstico. Pero esa visión idílica está lejos de ser una realidad. En la actualidad con lo que nos encontramos en nuestra pantalla, cuando nos sentamos ante ella dispuestos a buscar información, es con una serie de buscadores como Altavista, Excite, HotBot, Infoseek, Lycos, Yahoo, y tantos otros, que nos ponen en contacto con millones de páginas, con millones de documentos, pero mediante procedimientos que, si no andamos con mucho cuidado, nos inundan de información irrelevante, perdiéndose entre ellos muchos documentos que contienen información de gran valor, aunque no posean la estructura de los documentos impresos tradicionales.

Estamos, pues, con respecto a este tipo de materiales, en una fase muy primitiva en la que no existen normas para su catalogación, clasificación y localización; conceptos ellos mismos que requieren una profunda revisión en las actuales situación y circunstancias. Es aquí precisamente donde aparece la necesidad de meter algo de orden en el actual caos, donde aparece la necesidad de crear normas que regulen la descripción y la localización de los documentos electrónicos. Creación que tropieza con grandes dificultades si se tiene en cuenta la gran facilidad que existe para modificar los documentos (tanto los contenidos, como los autores, como los propietarios de derechos, etc), y también de modificar su localización al ser sumamente fluida la transferencia de los documentos digitales de unos depósitos a otros.

Por eso, para definir la Biblioteca Digital en un sentido más *estricto*, deben abordarse sistemáticamente cuestiones conceptuales y elaborarse normas, protocolos y procedimientos en las siguientes áreas:

1. Captura, clasificación, almacenamiento y organización de la información. (Interpretación de textos, del habla, de imágenes)
2. Sistemas de búsqueda, barrido, filtrado, resúmenes y visualización de la información. (Teorías sobre modelos, procesos inteligentes, aprendizaje; desarrollo de sistemas sobre simulación, navegación, optimización)
3. Redes, y usos de la información en la red. (seguridad, compresión, modelización, derechos de la propiedad intelectual)

En este sentido hace ya unos años que se está trabajando en Estados Unidos, en Europa y en Japón. En Estados Unidos se comenzó con una serie de proyectos de investigación, financiados por la NSF, ARPA y NASA, y encabezados por las Universidades de Carnegie-Mellon, Illinois, Michigan, Berkeley, Santa Barbara y Stanford, y en los que

colaboraron también grande empresas de informática, editoriales, etc. como Microsoft, IBM, Apple, Elsevier, Mc Graw Hill, O'Reilly, Enciclopedia Biritanica, BBC, Kodak, Xeros, entre otras. En Europa hay también un amplio plan llamado "Telemática para Bibliotecas", mediante el cual la Dirección General XIII de la Unión Europea, subvenciona proyectos de investigación en temas de interés para el desarrollo de la Biblioteca Digital.

4.2.- Objetos digitales y metadatos.

Los elementos primarios que conforman la biblioteca digital son los *documentos digitales*, que corresponden a una gran diversidad de materiales informativos; dada la diversidad de formatos y características de los materiales electrónicos, parece adecuado sustituir la denominación de *documento* (que tiene muchas connotaciones arcaicas) *digital o electrónico*, por el mas general de *objeto digital* en el que queda incluido una cantidad mas amplia de tipos de información digitalizada.

Por una parte, un *objeto digital*, puede ser considerado simplemente, como una secuencia de bits, de ceros y unos, ya que con ellos y una adecuada codificación se puede representar cualquiera de los materiales tradicionales. Pero, por otra, lo esencial de un *objeto digital*, es el contenido presentado al usuario, y otra serie de información adicional que sirve para indicar diversas propiedades de dichos contenidos (como los derechos de autor, o la forma de codificación, etc.) o que sirve para su identificación y localización, por tanto, podemos diferenciar dos partes en un objeto electrónico: su *contenido* y sus *metadatos*. Es decir, podemos considerar a un objeto digital compuesto de la siguiente forma:

- Contenido es la información creada por los autores, bien en forma digital, o bien convertida a este formato por la digitalización de textos, imágenes, sonidos, etc. Puede considerarse simplemente como una secuencia de bits, aunque en los casos prácticos estas secuencias de bits tienen una estructura y una semántica. Los modelos o esquemas de datos se usan para dar una semántica a los objetos digitales. El estudio de esta área esta todavía desarrollándose.
- Metadatos es información adicional sobre el contenido que puede ser descompuesta en las siguientes categorías:
 - *Asidero* es un metadato que se usa para la identificación unívoca del objeto y es de una importancia capital para facilitar su gestión.
 - *Propiedades* es la parte de los metadatos utilizada para describir los objetos digitales tales como dar la descripción bibliográfica, indicar los derechos de autor o las condiciones de acceso, decir cuándo un objeto puede cambiarse y quién puede cambiarlo, suministrar información sobre los formatos en que están digitalizados los objetos, etc.
 - *Firma* es la parte utilizada para certificar que el objeto no ha sido alterado.
 - *Transacciones* es la información que nos indica como ha sido usado el objeto, por ejemplo nos da información sobre quién, cuándo y cómo se ha utilizado o distribuido el objeto.

Por otra parte debemos considerar que, en general en una Biblioteca Digital no se buscan objetos aislados, sino grupos de objetos relacionados entre sí. Por eso es importante poder tratar *conjuntos de objetos*, ya que cuando se busca un "informe", un "programa", una

"obra musical", etc. frecuentemente se obtienen conjuntos de objetos relacionados entre si. Por ejemplo, una biografía de Goya y una colección de imágenes de sus obras. Un conjunto de objetos pueden ser representados en la biblioteca digital como un *objeto compuesto*, formado por los objetos del conjunto, o como un *meta-objeto* en cuyo caso solo contiene los asideros correspondientes a los objetos del conjunto. También es posible la combinación de los dos casos anteriores.

Los *metadatos* son datos que nos dan los atributos con los que se describe un objeto digital. Normalmente soportan ciertas funciones como: localización, documentación, evaluación, selección, etc. Los metadatos no solo son claves para la búsqueda, sino también y sobre todo sirven para facilitar el uso de los recursos (por ejemplo indicando las condiciones técnicas o comerciales para que ese recurso pueda ser utilizado) y para la interoperabilidad entre dominios diferentes.

Muchas comunidades científicas han empleado un gran esfuerzo en definir la manera de describir, con la necesaria riqueza y precisión, objetos electrónicos utilizados por los especialistas en su área de conocimiento. Esta multitud de situaciones hace difícil definir un único formato que contemple todas las posibilidades.

A cada objeto electrónico se puede asociar una diversidad de tipos de metadatos. Unos serían similares a la información bibliográfica contenida en los catálogos de las bibliotecas. Otros, facilitarían información a una aplicación cliente sobre el formato o la localización de objeto electrónico. En la definición de los metadatos también se tiene en cuenta el perfil de los usuarios del objeto, o las características de los proveedores de información. Los metadatos pueden estar contenidos en el mismo documento (como ocurre con los contenidos en el comando META de HTML), o estar separados de ellos como ocurre con los ficheros de holdings. Todas estas consideraciones conducen a una gran variedad en la estructura de los metadatos, unas muy simples, o casi inexistentes, como ocurre con los metadatos obtenidos automáticamente por los buscadores Web a partir del texto completo, o muy complejas las estructuras de los metadatos que incluyen formatos tipo MARC.

Una vez definida una estructura se plantea la cuestión de quien introduce la información correspondiente a los metadatos de un determinado objeto electrónico. En principio esta tarea puede ser compartida por varios creadores de metadatos. Por una parte está el propio *autor* quien puede aportar información sobre los contenidos y sobre algunos aspectos técnicos (como por ejemplo formatos). Por otra parte están los *gestores de los depósitos de información* (quienes aportarán metadatos sobre localización, características de los depósitos y otras informaciones técnicas). También existen otros *agentes* que pueden agregar información sobre los objetos digitales (por ejemplo sobre los derechos de utilización, aspectos comerciales, etc.). Evidentemente, algunos metadatos pueden y deben ser modificados en el tiempo (por ejemplo la fecha de última modificación de los contenidos), y otros deberán ser fijos e inalterables (por ejemplo los correspondientes a los números únicos de identificación universal). En unos casos los metadatos se crean de forma manual, en otros de forma automática mediante motores especiales, o de forma mixta.

Siguiendo la clasificación hecha por Demsey y Heery, podemos agrupar las estructuras de los metadatos en tres grupos

Grupo 1: formato simple, sin normalizar, desarrollado por la empresa que lo utiliza, a partir una indización del texto completo. En este grupo se pueden incluir los metadatos empleados por Lycos, Altavista, Yahoo, etc.

Grupo 2: Formato estructurado, usando normas que están en elaboración, empleando registros con campos y subcampos. Algo de los metadatos de este tipo son los definidos por Dublin Core, las plantillas IAFA, SOIF, etc.

Grupo 3: Formatos complejos, que siguen normas internacionales establecidas, que utilizan un sistemas de etiquetas muy elaborado. En esta categoría podemos incluir a MARC, EAD (Encode Archive Description), TEI , CIMI (Compter Interchange of Musical Information), etc.

Para ver algunos detalles, tomemos un ejemplo del *Grupo 2*: el Dublín Core.

El núcleo de Dublín

Ya hemos visto que existen una gran variedad de normas para definir la estructura de los metadatos. Algunas se han desarrollado para describir y proporcionar el acceso a un tipo particular de información, como por ejemplo la norma FGDC (Federal Geographical Data Committee de los Estados Unidos) para datos geoespaciales, o para atender formatos especiales de información como MARC. El *núcleo de Dublín* se ha desarrollado para proporcionar el acceso normalizado a una gama amplia de tipos diferentes de recursos de información y para permitir recuperarlos mediante un solo método de búsqueda.

Antes de examinar las peculiaridades del conjunto de elementos del *núcleo de Dublín*, veamos sucintamente una arquitectura desarrollada para gestionar una variedad de conjuntos de metadatos, conocida como marco de referencia de Warwick, (definido en esta Universidad en 1996). Que usa el modelo "recipiente-paquete", y proporciona un armazon conceptual, un método para el manejo de los *recipientes* y *paquetes* proporcionados por cualquier aplicación particular del modelo.

Un *recipiente* es, simplemente, cualquier mecanismo por agrupar paquetes. Un *paquete* puede ser de tres tipos: *paquete primario o directo* es el que contiene una o más piezas de metadatos; cada paquete primitivo será de un tipo (por ejemplo, un paquete de MARC, del *núcleo de Dublín*, o de FGDC); un *paquete indirecto* refiere a otros recurso de información (por ejemplo a través de un eslabón a su URL); *paquete recipiente* el es el que se puede formar de forma recursiva incluyendo *recipientes* entre sus elementos.

Esta estructura parece extremadamente simple, pero de hecho es un modelo poderoso para manejar metadatos. Sus ventajas son: la de ser modular (los metadatos se agrupan en paquetes); ser ampliable (no hay limitación en los tipos de paquete que pueden agregarse en un *recipiente*); ser distribuido (a través del uso de paquetes indirectos) y ser recursivo (ya que en un paquete puede también haber un recipiente)

En este marco nació el *núcleo de Dublín*, durante los años 1995 y 1996, para dar respuesta a la necesidad de mejorar la recuperación de recursos de información distribuidos por WWW, y se está desarrollando como una norma para metadatos genéricos usados en bibliotecas, archivos, gobierno y otros editores de información online

Aunque originalmente se pensó esta norma con unos objetivos bastantes reducidos, limitados a describir documentos tales como páginas de HTML, archivos PDF, o imágenes gráficas, en la práctica ha sido difícil definir cual es su alcance, y especialmente saber que cosas no incluye

Es una norma descriptiva, y no evaluativa, ya que no contempla el mantenimiento de sistemas de evaluación.

La norma del *núcleo de Dublín* se limitó deliberadamente a considerar un pequeño conjunto de elementos que fuesen pertinentes en una gama amplia de tipos de recurso de información. Actualmente son quince los elementos contenidos en la norma.

Quienes están implementando la norma han planteado varias cuestiones relativas a la semántica de los metadatos (reglas para definir el contenido o significado de los quince campos) y a la sintaxis (reglas para estructurar sus campos). En ambos casos la norma es inestable, y está sujeta a cambios frecuentes.

El desarrollo del *núcleo de Dublín* lo realiza un grupo informal de bibliotecarios y tecnólogos de la información, es de funcionamiento internacional, y sus actividades están auspiciadas por la OCLC en Dublín, Ohio. Trabaja de forma análoga a como lo hace el grupo ZIG, mediante reuniones periódicas de sus miembros y usando una lista de discusión (foro) de forma muy activa en Internet. El *núcleo de Dublín* no es todavía una norma reconocida por ISO (International Organization for Standardization) ni por IETF (Internet Engineering Task Force), aunque ha presentado un cierto número de documentos para su aceptación en los que se describe la semántica de los quince elementos del *núcleo de Dublín*, se da un método para codificar los metadatos del *núcleo de Dublín* en HTML

En la Home Page del *núcleo de Dublín* encontramos las siguientes definiciones de los quince elementos a la norma:

TITULO nombre dado al recurso por el CREADOR o EDITOR.

CREADOR personas u organizaciones responsables del contenido intelectual del recurso.

MATERIA tema del recurso, o palabras clave o frases que describen la materia o el contenido del recurso.

DESCRIPCION descripción textual del contenido del recurso, incluye los resúmenes en el caso de documentos o las descripciones de contenido en el caso de recursos visuales.

EDITOR entidad responsable de hacer disponible el recurso en su forma presente, como un editor, un departamento universitario, o una entidad corporativa.

COLABORADORES: personas u organizaciones, distintas de las especificadas en el elemento del CREADOR, que han hecho contribuciones intelectuales significativas al recurso pero cuya contribución es secundaria respecto a los individuos o entidades especificados en el elemento CREADOR.

FECHA: fecha en que el recurso se puso en circulación en su forma actual.

TIPO: categoría del recurso, como por ejemplo: texto, imagen, sonido, pagina web, novela, poema, etc.. Este es un elemento todavía no muy bien definido.

FORMATO: manera de representar los datos del recurso

IDENTIFICADOR: cadena de símbolos o número que identifican al recurso de forma unívoca.

FUENTE: trabajo, impreso o electrónico, del que se deriva el recurso.

IDIOMA: Lenguas del contenido intelectual del recurso.

RELACION: relación con otros recursos. Este elemento todavía no esta bien definido.

FONDOS: característica de las localizaciones espaciales y de la duraciones temporal del recurso.

DERECHOS: enlace con lugares en donde se de noticia de los derechos de propiedad intelectual.

5. Conclusión

Para terminar solo queremos resaltar el gran esfuerzo que se está haciendo en nuestros días por construir ese ideal de Biblioteca Digital, de Biblioteca Global, del que hemos hablado mas arriba, y que en esta construcción esta colaborando, y confluyendo en sus resultados, bibliotecarios (modificando sus antiguos elementos para adecuarlos a la gran eficacia de las nuevas tecnologías), con informáticos y tecnólogos de la información (incluyendo en sus procedimientos toda la sabiduría acumulada por los bibliotecarios con siglos de experiencia en la organización y gestión de la información). En la bibliografía que sigue pueden encontrar algunas referencias útiles para quienes deseen iniciarse en estas tareas.

6. Bibliografía.

6.1. Referencias

Barber, David .- Building a digital library : concepts and issues.- En: Library Technology Reports, September-October 1996.

Brown, David J. .- Electronic Publishing and Libraries: Planning for the Impact and Growth to 2003.- Compilado por David J. Brown.- Londres : Bowker Saur, 1996

Buckland, Michael K. + Lynch Clifford A.- "The Linked Systems Protocol and the future of bibliographic networks and systems." Information technology and libraries. 6(2):83-88 (1987).

Comisión Europea.- Developpement stratégiques de l'edition européenne à l'horizon de l'an 2000 : la challenge multimedia de l'Europe : raport.- Bruselas; Luxemburgo : Comisión Europea DG XIII-E, 1996

Denenberg, Ray.- "Data communications and OSI." Library hi tech. 8(4):15-32 (1990).

Denenberg, Ray.- Structuring and Indexing the Internet. Library of Congress, December 1996 Presentado en el Workshop on Earth Observation Catalogue Interoperability, patrocinado por la European Space Agency y el Centre for Earth Observation de la Comunidad Europea, 14-15 November 1996, en Ispra, Varese, Italia.

Denenberg, Ray.- "Open systems interconnection." En: Encyclopedia of library and information science, editado por Allen Kent, vol. 44, suplemento 9, 210-233. New York: Marcel Dekker, 1989.

Fenly, Judith G. + and Beacher Wiggins.- The Linked Systems Project: a networking tools for libraries. Dublin (Ohio): OCLC, 1988.

Fundesco.- Multimedia 1996 : tendencias.- Madrid : Fundesco, 1996

García Camarero, Ernesto.- El mundo de la Informática. Cuadernos para el Diálogo. XXVII extraordinario, p. 78-82. Octubre 1971.

Hammer, Sebastian + John Favaro.- Z39.50 and the World Wide Web. «D-Lib magazine», March 1996.

Lang, Brian.- The Legal Deposit of Electronic Publications / Working Group of the Conference of Directors of National Libraires (CDNL) Presidida por Brian Lang...- 1997.-

Lynch, Clifford.- Using the Z39.50 information retrieval protocol in the Internet Environment. IETF/IIIR Working group Internet draft.

Lynch, Clifford A.- "Access technology for network information resources." CAUSE/EFFECT. 13(2):15-20 (1990).

Lynch, Clifford A.- "Interoperability: the standards challenge for the 1990s." Wilson library bulletin. 67(7):38-42 (1993).

Lynch, Clifford A.- "The client-server model in information retrieval." En: Interfaces for information retrieval and online systems, editado por Martin Dillon. Greenwood Press, 1991, 301-318.

Mackenzie Owen, + J.S. J.V.d. Wale.- Deposit Collections of electronic publications.- Bruselas; Luxemburgo : Comisión Europea DG XIII-E, 1996.

McCallum, Sally H.- "Information technology standards: implementation, maintenance, and coordination." Wilson library bulletin. 67(7):43-45 (1993).

Millán, José Antonio.- La edición electrónica y multimedia.- Madrid : Federación de Gremio de Editores de España, 1996

Oliván Plazaola, Montserrat + Luis Ángel García Melero.- Publicaciones electrónicas y depósito legal.- En: II Xornadas de Arquivos, Bibliotecas e Museos de Galicia : Cooperación: realidade e futuro : A Coruña, 24-26 de abril de 1997.- Santiago de Compostela: Xunta de Galicia, 1997.-p. 325-337

Peterson Bishop, A. .- "Working toward an understanding of digital library user: a report on the user research efforts of the NSF/ARPA/NASA DLI projects," D-Lib Magazine, October 1995

Roumieux, Olivier.- L'impact de l'Internet sur la profession de bibliothécaire. - Ecole Polytechnique, Mémoire de fin d'études, 1996.-

Samain, Thierry.- L'accès aux catalogues des bibliothèques par Internet. - Mémoire d'étude ENSSIB. 1996.

Scolari, Antonio.- Antonio Scolari. Gli standard OSI per le biblioteche: dalla biblioteca-catalogo alla biblioteca-nodo di rete. Milán: Editrice Bibliografica, 1995.

Turner, Fay.- An Overview of the Z39.50 Information Retrieval Standard.- National Library of Canada (fay.turner@nlc-bnc.ca), Julio, 1995. Revisiones: Junio, 1996; enero, 1997.

Van House, N. A. et. al. .- "User centered iterative design for digital libraries: the Cypress experience," En: D-Lib Magazine, February 1996

Normas relacionadas con la Z39.50

- ANSI/NISO Z39.58-1992 -- Common Command Language for Online Interactive Information Retrieval
- ISO 2709 -- Documentation - Format for Bibliographic information interchange on Magnetic Tape 1981.
- ISO 4217 - Codes for the representation of currencies and funds 1990.
- ISO 7498 -- Information Processing Systems - Open Systems Interconnection - Basic Reference Model 1984.
- ISO 8649 -- Information Processing System - Open Systems Interconnection - Service Definition for the Association Control Service Element 1987.
- ISO 8650 -- Information Processing Systems - Open Systems Interconnection - Protocol Specification for the Association Control Service Element 1987.
- ISO 8777 -- Information and Documentation - Commands for Interactive Text Searching.
- ISO 8822 -- Information Processing Systems - Open Systems Interconnection - Connection Oriented Presentation Service Definition 1988.
- ISO 8824 -- Information Processing Systems - Open Systems Interconnection - Specification of Abstract Syntax Notation One (ASN. 1) 1990.
- ISO 8825 -- Information Processing Systems - Open Systems Interconnection - Specification of Basic Encoding Rules for Abstract Syntax Notation One (ASN. 1) 1990.
- ISO 10160 -- Information and Documentation - Interlibrary Loan Application Service Definition for Open Systems Interconnection 1991.
- ISO 10161 -- Information and Documentation - Interlibrary Loan Application Protocol Specification for Open Systems Interconnection 1991.
- ISO 10163 -- Information and Documentation - Search and Retrieve Application Protocol Specification for Open Systems Interconnection 1991.
- ISO -- International Register of Coded Character Sets To Be Used with Escape Sequences 1992.
- Library of Congress -- USMARC Code List for Languages 1989.