

▲ Tabla de contenido

RECUPERACIÓN DE INFORMACIÓN EN CLASE: UN ENFOQUE BASADO EN EL CONCEPTO DE METADATOS

MARCO ANTONIO MONTES FLORES

Centro Regional ASFIS-DGB

Departamento de Bibliografía Latinoamericana

Dirección General de Bibliotecas

Universidad Nacional Autónoma de México (UNAM)

Resumen

La inclusión de la base de datos CLASE dentro de la World Wide Web, sin lugar a dudas representó un gran avance para su difusión y alcance. Sin embargo, la adecuación de CLASE dentro de un formato electrónico, trajo consigo la pérdida de visibilidad de su variedad temática, la cual abarca 17 disciplinas que van desde la administración y la contaduría hasta la religión, pasando por el arte, el derecho, la política, etc. La clasificación temática de CLASE fue diseñada originalmente para la versión impresa de la base, ahora, al adecuarse para su consulta en línea, esta diversidad no está manifiesta expresamente en el actual formato de CLASE.

Por lo tanto, con el afán de hacer visible la diversidad temática de CLASE y de facilitar una recuperación mucho más precisa de la información por parte de los usuarios, se propone una nueva forma de presentación de la página electrónica de CLASE basada en el concepto de metadatos, se analiza un posible diseño estructural de los enlaces dentro de la base de datos y se ofrece un ejemplo de consulta.

Palabras clave: *bases de datos, CLASE, metadatos, recuperación de información*

Introducción

Sin lugar a dudas el impacto que Internet ha tenido en nuestro quehacer cotidiano es definitivo, puede observarse su presencia en las más variadas ramas de la actividad humana desde la ciencia, la educación, el comercio, etc. hasta la forma en que nos comunicamos diariamente. Esta revolución necesariamente afectó de manera primordial, y seguirá afectando, la forma de generar, distribuir y recuperar la información.

En los años recientes hemos visto como la comunicación impresa se ha ido transformando para adoptar formatos electrónicos que han permitido alcanzar una serie de ventajas en cuanto a la facilidad de producción, la ampliación de su cobertura y en algunos casos en la disminución de costos; así los documentos electrónicos se han convertido en una alternativa real para el intercambio de información. Todo esto ha propiciado una verdadera explosión en cuanto al crecimiento de las publicaciones digitales y a un desarrollo superlativo de la información a la cual tenemos alcance; paradójicamente ante tan abrumadora cantidad de información, (textos, imágenes, vídeo, audio, etc.) nos resulta cada vez más difícil acceder a la información precisa que requerimos, basta navegar un poco por Internet o utilizar en su forma más simple un motor de búsqueda para darnos cuenta de la cantidad de información y la poca precisión que encontramos

en la mayoría de los casos, lo anterior en lugar de propiciar un mayor uso de los recursos de información digitalizados, llega a provocar actitudes de rechazo por parte de los usuarios.

Por todo esto, podemos ver que nos enfrentamos al problema de establecer y desarrollar sistemas que permitan una adecuada organización de los recursos de información electrónica con el fin de que estos puedan ser identificados, localizados y recuperados de manera efectiva.

Para empezar debemos reconocer las características que dificultan el manejo de estos recursos electrónicos, dentro de ellas cabe mencionar las siguientes:

- Variabilidad de presentación. La configuración de los navegadores locales o el estilo de las páginas determinan la forma en que aparece un documento, esto en algunas ocasiones es generado de manera aleatoria.
- Relaciones mal definidas con otras formas. Al menos una parte de las publicaciones electrónicas se deriva de publicaciones en otros medios (papel, radio, televisión), etc.), en ocasiones no es posible determinar el grado de integridad del contenido del recurso electrónico con respecto a su fuente original.
- Mutabilidad y movilidad. Salvo en el caso del material generado localmente, no hay una certeza total de que los materiales pudiesen sufrir una modificación. Por otro lado la existencia de las páginas electrónicas es, por lo general, temporal, ya que están sujetas a cambios de ubicación, modificación de contenido, etc.

Toda esta dinámica de la publicación digital altera los modelos establecidos para su control, por otro lado, podemos observar que las formas básicas de las publicaciones electrónicas aún no son estables, continúan evolucionando de una manera difícilmente predecible.

Ante toda esta corriente de cambio surge la necesidad de analizar, rediseñar y ampliar los procesos y estándares existentes, se necesitan revalorar las prácticas de catalogación tomando en cuenta las nuevas formas de publicación las cuales requieren de actividades de administración de colecciones no aplicables en otros medios, así como de medios de acceso innovadores que maximicen su valor y uso.

Dentro de toda esta dinámica surge entre los profesionales de la información la palabra metadatos, cuyo concepto es importante tanto para los autores como para los recuperadores de información electrónica, los metadatos, usados de manera adecuada, ponen a disposición de los usuarios la información al marcar su contenido de manera consistente. Los metadatos marcan el camino a seguir para encontrar la información requerida.

¿Qué son los metadatos? Aún cuando existen diversas definiciones de esta palabra, la más clara, desde mi punto de vista, es la que define a los metadatos como datos acerca de datos. Esto es, que un elemento de metadatos describe un recurso de información, o ayuda a facilitar el acceso a este recurso. Por ejemplo, una ficha catalográfica de una biblioteca es una colección de metadatos que están unidos a un libro mediante el número de registro de este libro. El término metadatos se aplica de manera general a recursos electrónicos y se refiere a datos en su sentido más amplio, p. ej. colecciones de datos, información textual, gráficas, música, y prácticamente a todo lo que sea

sujeto de aparecer de manera digitalizada, aún cuando debemos recordar que toda la labor de catalogación e indización realizada por la comunidad bibliotecaria durante décadas también puede ser descrita como metadatos.

De hecho, todas las razones por las cuales la catalogación y el indizado son necesarios para manejar los recursos impresos se aplican aún más enfáticamente a los metadatos para los documentos electrónicos. Es prácticamente imposible navegar por los millones de documentos ubicados en miles de servidores diferentes alrededor del mundo sin utilizar algún tipo de ayuda.

Por todo lo anterior concluimos que el propósito básico de los metadatos es facilitar y mejorar la recuperación de información, de ahí su importancia crucial en la búsqueda de información. Dado que en la actualidad las búsquedas de información se basan principalmente en la comparación de palabras clave con palabras en el texto de los documentos, cualquier proceso que simplifique el proceso de comparación o aumente su grado de estandarización mejorará los métodos de búsqueda de información; es básicamente en el proceso de estandarización de los metadatos donde se da la mayor discusión y problemática para lograr su consolidación.

Para que un esquema de metadatos pueda ser exitoso, aparte de su función descriptiva también debe de establecer una estructura y terminología estándar y es aquí en donde los esfuerzos de los profesionales de la información están enfocados. Así podemos ver que existen diferentes grupos de trabajo con filosofías distintas en cuanto a la estructura y al contenido de los esquemas de metadatos, cabe mencionar aquí al Núcleo de Dublín, USMARC, TEI (Text Encoding Initiative), GILS (*Government Information Locator Service*).

Al revisar la metodología de los diferentes proyectos nos encontramos con dos enfoques distintos de abordar el problema: por un lado existe una tendencia a organizar sólo la estructura de los datos y no su contenido; dicho de otra manera sólo se ocupan de los campos que debe tener una estructura de metadatos y no les preocupa mucho la manera en que los datos son codificados dentro de esta estructura produciendo así una catalogación por demás sencilla, aunque muy general y con pocas posibilidades de precisión con respecto a la búsqueda de información. La tendencia alterna considera no sólo la estructura de los datos sino también la forma en que se codifican éstos y propone el uso de ciertas normas, tales como el uso de vocabularios controlados, thesauros, etc. Aquí, la catalogación es más compleja pero la riqueza en cuanto a la precisión en la recuperación de información es superior. Dentro de estas dos tendencias existe toda una gama de enfoques que tratan de conciliar ambas corrientes. Finalmente, existe un tercer enfoque el cual consiste en crear esquemas específicos de acuerdo al tipo de información que se maneje.

El caso de CLASE

Durante sus veinticinco años de existencia la base de datos CLASE se ha distinguido tanto por su diversidad temática, como por su cobertura regional, lo que hacen que se constituya en una base de datos única en América Latina. Durante todo este tiempo ha sido modificada con el propósito de adecuarse a las innovaciones tecnológicas que han provocado verdaderas revoluciones en los medios de comunicación, así hemos visto que CLASE ha tenido diversas formas de presentación, inicialmente como un índice impreso, posteriormente estuvo disponible su consulta en línea utilizando el software Minisis, a esto siguió su presentación en CD-ROM y finalmente su

incorporación a Internet a través de la página electrónica de la Dirección General de Bibliotecas (DGB).

La estructura original de CLASE presentaba las referencias agrupadas en disciplinas y subdisciplinas, cabe mencionar que CLASE cubre 17 disciplinas, estas son: Administración y contaduría, Antropología, Arte, Bibliotecología y ciencia de la información, Ciencias de la comunicación, Demografía, Derecho, Economía, Educación, Filosofía, Historia, Lingüística y literatura, Política, Psicología, Relaciones internacionales, Religión y Sociología; aparte de incluir listados de palabras clave, de autores y de revistas.

Esta estructura se respetó tanto para la consulta en línea, como para el CD-ROM; sin embargo al adecuar la base de datos para su consulta vía Internet, se cambió esta estructura y la diversidad temática de la base de datos quedó oculta, ni siquiera en la explicación de la base se da a conocer su temática, este detalle que pudiera parecer poco significativo tiene una gran repercusión en cuanto a la efectividad de las búsquedas.

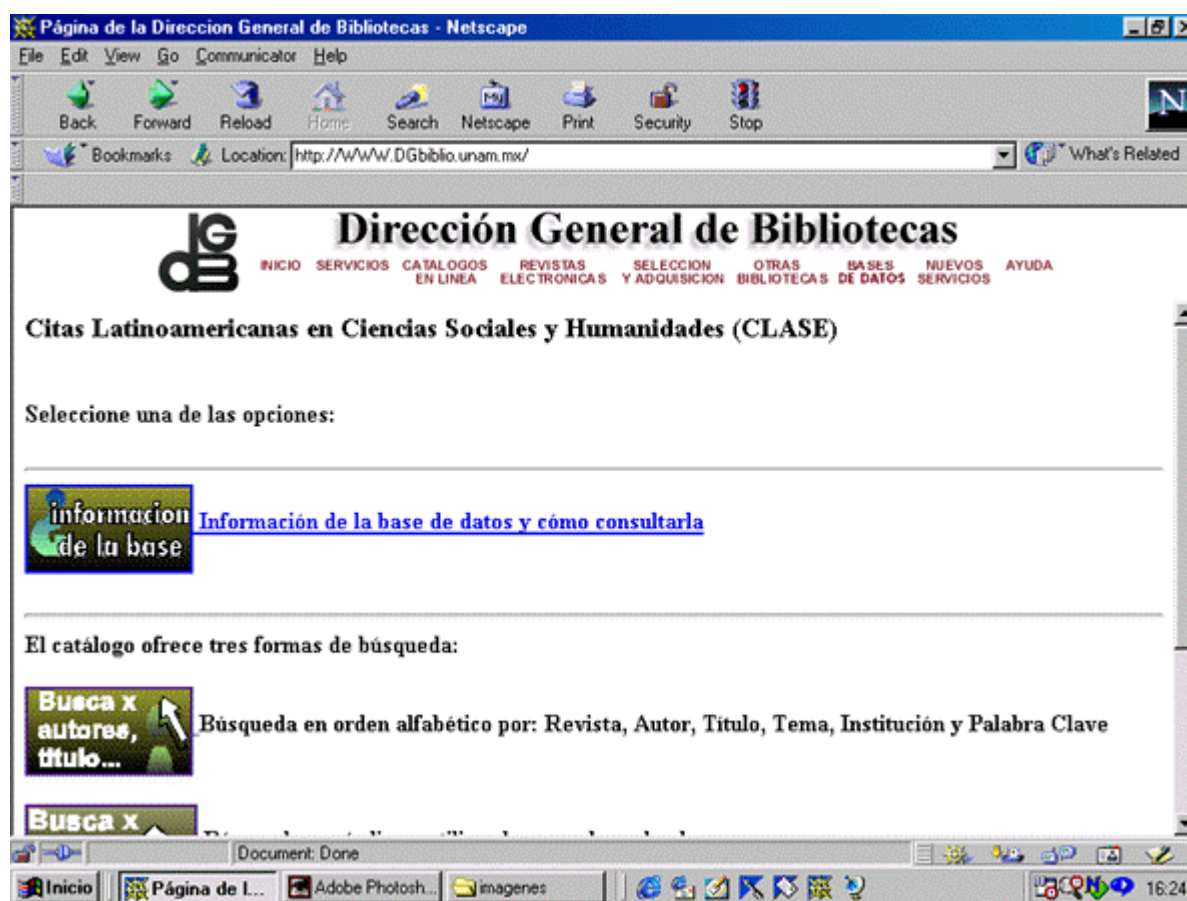


Figura 1. Página inicial de CLASE

Cuando un usuario nuevo consulta la base, bien poco puede deducir del contenido de esta con sólo ver su nombre y no obtendrá mucha mayor información al leer la descripción de la base de datos (sí llegase a leerla). Al utilizar los recursos de búsqueda (por palabras, por índices, por

autores, título, etc.) la cantidad de referencias obtenidas puede ser muy grande pero la precisión es poca (Figura 1). Así vemos que al buscar la palabra sexualidad se recupera una cantidad enorme de referencias con una diversidad temática muy grande que va desde el cuento hasta un artículo científico (Figura 2).

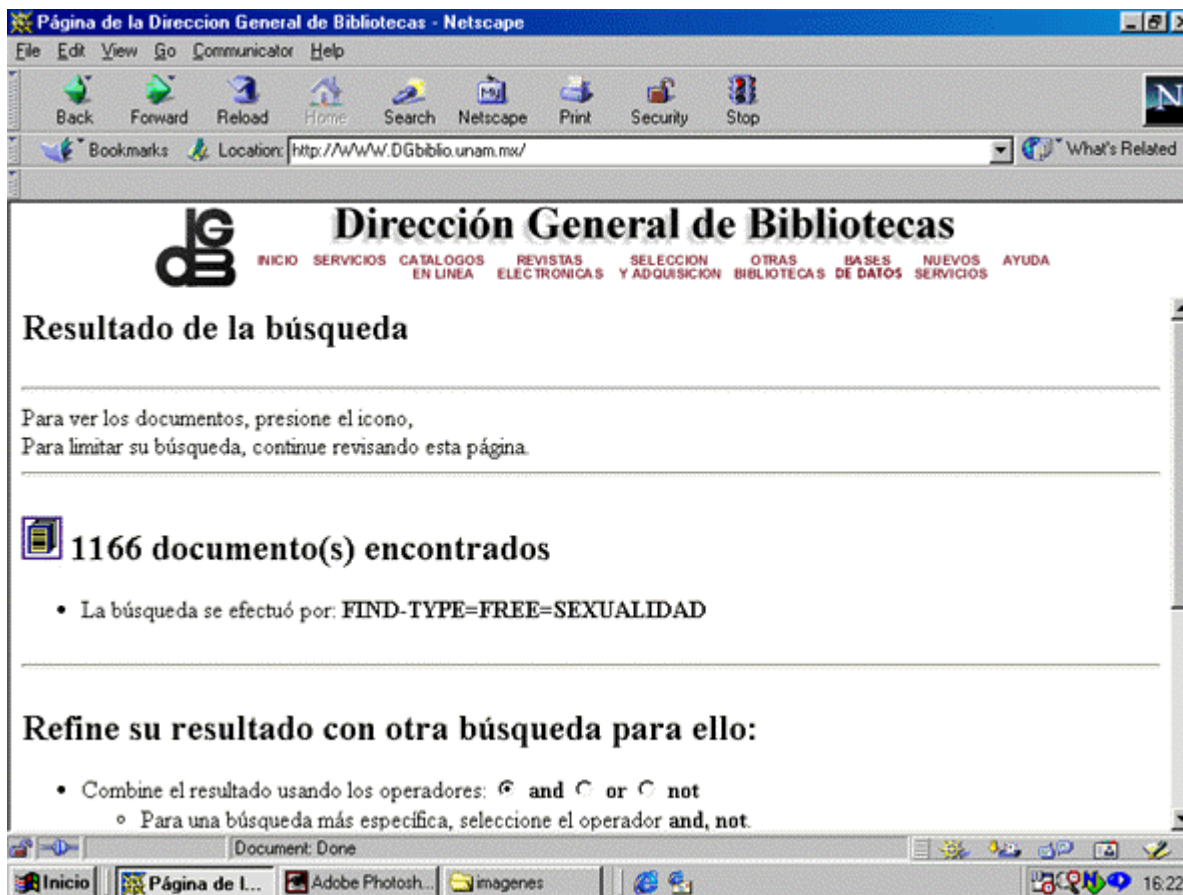


Figura 2. Resultados de búsqueda de la palabra sexualidad

Para un recuperador de información experimentado no será difícil refinar su búsqueda, pero el espectro de nuestros usuarios es mucho más amplio, por lo tanto si queremos mejorar la precisión en la recuperación de las búsquedas en nuestra base de datos debemos hacer algo.

Es aquí donde la utilización de los esquemas y estructura de los metadatos puede ser útil. En este caso proponemos adecuar sólo los esquemas al ambiente de Internet puesto que la organización original de la base de datos ya tiene la estructura básica de los metadatos, esto es, la base incluye disciplinas que agrupan cada una a una serie de subdisciplinas a las cuales van asociadas las referencias; lo anterior junto con todo el trabajo ya realizado de normalización de la base, no sólo de los temas sino también de las palabras clave, de la forma de inclusión de los autores, etc. nos ofrece una muy buena oportunidad para elaborar una vía de búsqueda con la consistencia necesaria para lograr una mayor precisión en la recuperación de la información.

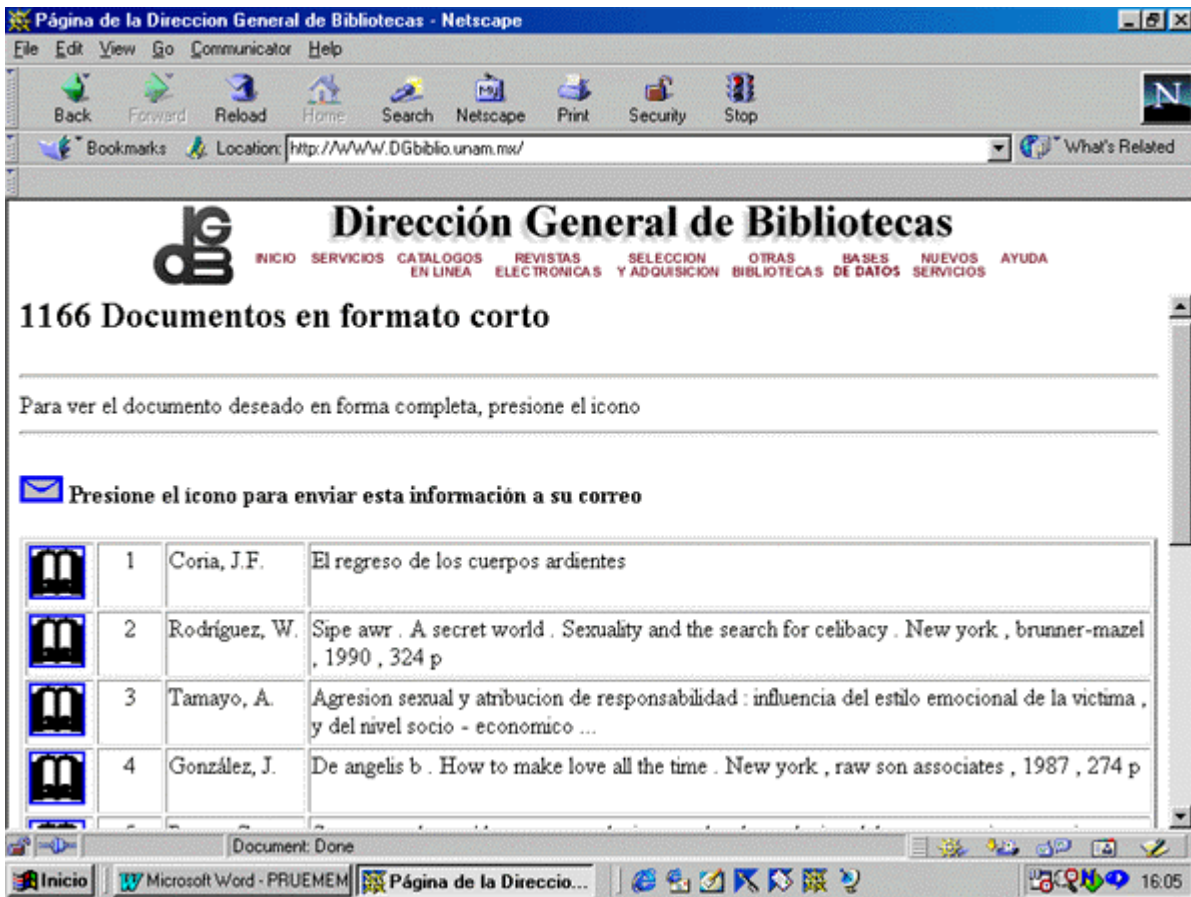


Figura 3. Presentación de resultados

Así tenemos que la clasificación temática de CLASE nos proporcionará los datos necesarios para orientar al usuario acerca de lo que puede encontrar en la base; esto constituye la primera fase en la búsqueda de información, la cual ubicará al usuario dentro de los 17 temas que tiene CLASE y a partir de aquí estará en posibilidad de elegir el tema de su interés enfocando su búsqueda con mayor precisión. Una vez seleccionado el tema aparecerá una variedad de subdisciplinas relacionadas con éste y por último el usuario podrá buscar por la palabra o palabras que sean de su interés.

¿Cuál es la diferencia con respecto al esquema actual de búsqueda? La diferencia básica radica en que esta ya no se realiza en todo el universo de la base, sino que ahora al usar la estructura de los metadatos el usuario va enfocando de manera secuencial su investigación y los datos que obtendrá serán sólo los que estén relacionados con su tema ya que el motor de búsqueda sólo recuperaría las referencias asociadas con el tema elegido.

Técnicamente, esta vía de búsqueda no implica nada más que la creación de los archivos correspondientes a cada disciplina así como de los enlaces de las subdisciplinas con éstas y por último ligar las referencias correspondientes a cada disciplina con su respectivo archivo; prácticamente todo el trabajo está hecho, puesto que las subdisciplinas forman parte de todos los

registros de CLASE y ligarlas a una disciplina implica un solo cambio en la ruta de enlace de los registros, incluso el motor de búsqueda que se usa actualmente no necesita cambiarse.

Ahora bien, retomando el ejemplo de la búsqueda de la palabra sexualidad, ahora bajo el esquema de los metadatos, podremos ver que todo el universo de 1166 referencias se ha fragmentado de acuerdo a la disciplina y subdisciplina a la que hayan sido ligadas, obviamente la cantidad de referencias recuperadas será menor pero tendrán una relevancia y precisión mucho mayor que con el esquema anterior, gracias a los filtros por los que ha sido sometido el proceso de búsqueda. (Figura 3)

Es un hecho que el desarrollo de las tecnologías de información avanza de manera impresionante y que la presión que ejerce sobre los proveedores de información para cambiar sus formatos de presentación es muy grande, pero la necesidad de adecuarse a las nuevas tecnologías no debe implicar pérdidas en el contenido y calidad de las fuentes de información, sino por el contrario todos los esfuerzos deben de encaminarse a mejorar la calidad del acceso y recuperación de la información.

Citas Latinoamericanas en Ciencias Sociales y Humanidades

C L A S E

Disciplinas

ADMINISTRACIÓN Y CONTADURÍA
ANTROPOLOGÍA
ARTE
BIBLIOTECOLOGÍA Y CIENCIA DE LA INFORMACIÓN
CIENCIAS DE LA COMUNICACIÓN
DEMOGRAFÍA
DERECHO
ECONOMÍA
EDUCACIÓN
FILOSOFÍA
HISTORIA
LITERATURA Y LINGÜÍSTICA
POLÍTICA
PSICOLOGÍA
RELACIONES INTERNACIONALES
RELIGIÓN
SOCIOLOGÍA

SUBDISCIPLINAS

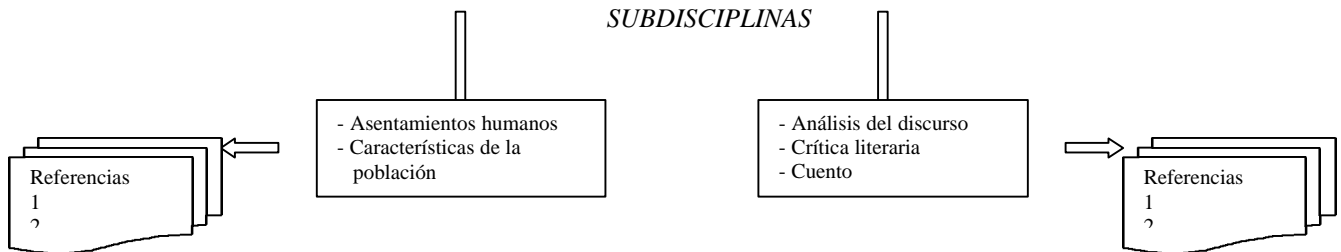


Figura 4. Esquema de recuperación de información.

Bibliografía

- Association for Library Collections & Technical Services (ALCTS) . *Task Force on Metadata. Summary Report.* <http://www.ala.org/alcts/organization/ccs/ccda/tf-meta3.html> (consulta: junio de 1999)
- Berners-Lee, Tim. *Metadata Architecture: Documents, metadata, and links.* <http://www.w3.org/DesignIssues/Metadata> (consulta: enero de 1998)
- Cathro, Warwick. *Metadata: An overview. A paper given at the Standards Australia Seminar.* <http://www.nla.gov.au/nla/staffpaper/cathro3.htm> (consulta: agosto de 1997)
- *Dublin Core metadata.* <http://www.purl.org/DC/> (consulta octubre de 2000)
- Gorman, Michael. ¿Metadatos o catalogación? Un cuestionamiento erróneo. En: *Internet, metadatos y acceso a la información en bibliotecas y redes en la era electrónica/* Filiberto Felipe Martínez Arellano, Lina Escalona Ríos, (Comps.) México:UNAM, Centro Universitario de Investigaciones Bibliotecológicas; Infoconsultores 2000: 1-20
- Library of Congress. *Metadata, Dublin Core and USMARC: A review of current efforts. MARBI Discussion paper no. 99, Library of Congress.* <gopher://marvel.loc.gov/00/.listarch/usmarc/dp99.doc> (consulta enero de 1997)
- Lynch, Clifford. *Identifiers and their role in networked information applications.* ARL: A Bimonthly Newsletter of Research Library Issues and Actions 194. <http://www.arl.org/newsltr/194/identifier.html> (octubre de 1997)
- McKiernan, Gerry. Puntos de vista: acceso y navegación convencionales y “neoconvencionales” en colecciones digitales. En: *Internet, metadatos y acceso a la información en bibliotecas y redes en la era electrónica/* Filiberto Felipe Martínez Arellano, Lina Escalona Ríos, (Comps.) México: UNAM, Centro Universitario de Investigaciones Bibliotecológicas; Infoconsultores 2000: 49-69