

artículos

La indización por materia de las monografías en los catálogos en línea de bibliotecas generales

Ana María Martínez

Licenciada en Bibliotecología y Documentación y profesora titular. Departamento de Bibliotecología de la Facultad de Humanidades y Ciencias de la Educación, Universidad Nacional de La Plata, 48 y 6-5° piso, 1900 La Plata, Argentina

Mariana Pichinini

Bibliotecaria Documentalista, adscrita a la Cátedra de Clasificación I. Departamento de Bibliotecología de la Facultad de Humanidades y Ciencias de la Educación, Universidad Nacional de La Plata, 48 y 6-5° piso, 1900 La Plata, Argentina

RESUMEN

En esta revisión se reseñan algunos problemas relacionados con la indización por materia de las monografías en los catálogos en línea de bibliotecas generales, que pueden resumirse en la escasa cantidad de términos descriptivos en cada registro bibliográfico, las incoherencias en la indización y las limitaciones de los lenguajes de indización. Se han propuesto tres soluciones: aumentar los puntos de acceso por materia en cada registro, combinando las palabras claves de la lengua natural con términos verbales de indización (encabezamientos de materia y descriptores) y sistemáticos (notaciones); facilitar el acceso en línea al catálogo de autoridades y al lenguaje de indización tanto para los indizadores como para los usuarios; y mejorar los lenguajes de indización. Esta última opción se ha visto afectada por las dificultades económicas que impiden o hacen muy difícil la creación de nuevos lenguajes de indización o la reestructuración de aquellos tradicionales como las listas de encabezamientos de materia y las clasificaciones de cimas, y más aún, la reclusión o reindización de las colecciones existentes. Las experiencias obtenidas en los países desarrollados pueden ayudar a los bibliotecarios latinoamericanos a discutir estos problemas, prevenir los mismos errores y brindar mejores catálogos en línea a los usuarios.

ABSTRACT

Subject indexing of monographs in general libraries on-line catalogs. Some problems related to subject indexing of monographs in general libraries on-line catalogs are reviewed. The problems include the small number of descriptive terms in bibliographical records, inconsistency of indexing, and obsolescence of indexing languages. Three solutions have been proposed: to increase the subject access points for each bibliographical record, combining keywords from natural language (title and abstract or content note) with verbal controlled terms (subject headings or descriptors), and notational controlled terms (classifications); to provide on-line access to authority files and indexing languages, both for indexers and users, and to improve the indexing languages. This last proposal has been affected by economic issues, making it very difficult to create or rearrange those traditional indexing languages like subject headings lists and de cimal classifications, as well as the reclassification or reindexing of library collections. The experiences obtained in developed countries may help Latin American librarians to discuss these problems, to prevent the same errors, and to offer better on-line catalogs for users.

El acceso por materia a la información es la característica que distingue al catálogo de biblioteca de un simple inventario y que lo convierte en un instrumento de recuperación de información y no sólo de datos.

Sin embargo, a pesar de los adelantos alcanzados por la tecnología, aún no se ha podido dar a los usuarios una respuesta satisfactoria en este aspecto, sobre todo

en lo que concierne a la indización por materia de las monografías en los catálogos en línea de bibliotecas generales.

Inicialmente, al computarizarse los catálogos, la decisión de los bibliotecarios fue muy sencilla, incluso inocente: pasar la ficha de 7.5 x 12.5 cm a la máquina, es decir, reciclar el catálogo tradicional en un ambiente electrónico (Bates, 1989; Culkin, 1989; Lancaster, 1991; Su,

1994). En opinión de Mary Dykstra (1989) muchas de las experiencias actuales son sólo **catálogos de fichas sobre ruedas**, que han arrastrado consigo lo que Klugman (1989) y Larson (1991a) denominan problemas endémicos de la indización.

Paradójicamente, la búsqueda por materia es la más solicitada por los usuarios —59% de todas las búsquedas— y la que

mayor índice de fracasos presenta —40% de respuestas con cero registros recuperados (Kern-Simirenko, 1983; Larson, 1991b; Peters, 1989).

Este alto índice de fracasos ha sido relacionado con la falta de coincidencia entre el vocabulario del usuario y el vocabulario del sistema. Alyson Carlyle (1989), de la University of California at Los Angeles, ha calculado que el promedio de coincidencia entre los términos del usuario y los encabezamientos de materia de la Library of Congress (LCSH) es de un 50%, mientras que Cousins (1992) de Gran Bretaña, obtuvo 62% de coincidencia para LCSH y 30% para la Clasificación Decimal de Dewey (CDD) en una biblioteca politécnica y 73% y 36% respectivamente en una biblioteca pública.

Esta falta de coincidencia ha sido atribuida a tres causas principales: pocos términos descriptivos en cada registro bibliográfico, incoherencias en la indización y limitaciones de los lenguajes de indización.

Como solución se ha propuesto aumentar los puntos de acceso por materia, facilitar el acceso en línea al catálogo de autoridades y al lenguaje de indización y mejorar estos últimos (Bates, 1989; Buckland, 1992; Larson, 1991b; Su, 1994).

LOS PUNTOS DE ACCESO POR MATERIA

El catálogo de una biblioteca general, lo mismo en fichas que en computadora, incluye principalmente monografías. Por lo general, la indización por materia se limita a un nivel monográfico, sin resumen y con escasos términos de indización. En los Estados Unidos se ha calculado que el promedio de encabezamientos de materia en los catálogos en línea de bibliotecas generales es de 1.8 por registro, con lo cual los usuarios no pueden lograr ni una adecuada exploración de los términos ni mucho menos una combinación con los operadores booleanos.

En 1977 Pauline A. Cochrane, de la Syracuse University de Estados Unidos, propuso aumentar los puntos de acceso por materia para los registros de monografías, incluyendo los términos de sumarios e índices analíticos en el campo de resumen, para ser recuperados por lengua natural. Con el fin de evaluar esta

propuesta, se comparó un catálogo de 1979 monografías en humanidades y ciencias sociales, cuya indización se hizo con formato MARC, por un lado, y con el **formato aumentado**, por el otro; esta última opción se denominó BOOKS y contaba con un promedio de 32.4 términos descriptivos de materia por registro. En 90 búsquedas de prueba la recuperación promedio con MARC fue de 56 registros en 8 minutos, mientras que con BOOKS fue de 130 registros en 4 minutos —233% de los registros MARC en 50% del tiempo (Settel y Cochrane, 1982).

A pesar de ciertas consideraciones económicas (Holley y Killheffer, 1982), la idea de aumentar los puntos de acceso por materia ha sido aceptada y no se ha limitado al uso de la lengua natural. Como es sabido, la combinación de la lengua natural con los lenguajes de indización mejora la calidad de la recuperación, ya que permite aumentar tanto la certeza (*recall*) como la precisión, aun que de diferente manera (Aitchison y Gilchrist, 1987; Dubois, 1987; Lancaster, 1991; Svenonius, 1986):

- a) La lengua natural aumenta la certeza cuando se incluyen resúmenes extensos y texto completo, por la exhaustividad de la cobertura y aumenta la precisión por el nivel de especificidad y actualización de sus términos.
- b) El lenguaje de indización aumenta la certeza mediante el control de sinónimos y las relaciones asociativas, mientras que aumenta la precisión por el uso de términos compuestos, el control de homónimos y polisemas y la utilización de vínculos y roles.

Shirley A. Cousins (1992), del University College of Wales de Gran Bretaña, tradujo los términos de sumarios e índices analíticos a PRECIS. Realizó tres pruebas con distintos resultados: formato MARC sin términos de sumarios e índices 13% de certeza; formato aumentado recuperado por lengua natural 76% de certeza; formato aumentado recuperado por PRECIS 81% de certeza. La diferencia de certeza entre el formato MARC y el formato aumentado es concluyente; la diferencia entre la lengua natural y PRECIS parece no justificar el costo de la traducción de los términos de sumarios e índices a un lenguaje de indización, pero es innegable que logró mejorar la certeza.

En forma similar, Gunnar Knutson (1991) llevó a cabo una investigación en la University of Illinois at Chicago con el propósito de establecer la relación entre el formato aumentado y el uso de la colección, evaluado mediante la circulación. Dividió su colección experimental en tres grupos: a) registro MARC aumentado con más encabezamientos de materia (promedio de 7.5 encabezamientos por registro) y una nota de contenido en lengua natural, b) registro MARC común, y c) registro MARC aumentado sólo con la nota de contenido. La circulación para el primer grupo fue de 46%, mientras que para los otros dos fue de 26% y 28% respectivamente.

También se han aumentado los puntos de acceso utilizando esquemas de clasificación como la Library of Congress Classification (LCC), la Dewey Decimal Classification (CDD) y la Clasificación Decimal Universal (CDU).

Ray R. Larson (1991a), de la University of California at Berkeley, ha desarrollado el sistema Cheshire que funciona de la siguiente manera: 1) los registros documentales o bibliográficos son copiados de un servicio de procesos técnicos; 2) la signatura topográfica (*call number*) asignada por la Library of Congress con LCC es reprocesada, eliminando la signatura librística (*Cutter number*) y conservando solamente la signatura de clase (*class number*), que de esta forma se convierte —automáticamente— en una notación de LCC; 3) el sistema consta de dos tipos de registros: el registro bibliográfico convencional con los datos del documento y el registro recolector (*cluster record*), que agrupa los títulos y los encabezamientos de materia de todos los registros bibliográficos que poseen una idéntica notación de LCC. En el momento de la búsqueda, cuando el usuario ingresa sus términos en lengua natural, el sistema le presenta en pantalla el registro recolector; entonces el usuario selecciona el título o el encabezamiento de materia y el sistema recupera y exhibe los registros bibliográficos correspondientes. Así, 30.471 registros bibliográficos que contenían 221.042 encabezamientos de materia fueron reunidos bajo 8.435 notaciones; Larson opina que esta modalidad permite suplir las relaciones equivalentes y asociativas de un tesoro.



Durante mucho tiempo se ha postulado que los lenguajes precoordinados no son aptos para la computadora. Sin embargo, la Reunión Satélite de IFLA sobre Indización, realizada en Lisboa, Portugal puso de manifiesto que los sistemas precoordinados han sido preferidos para los catálogos en línea de las bibliotecas generales en numerosos países.



El Dewey Decimal Classification Online Project fue una investigación realizada en los años ochenta y financiada por el Council of Library Resources, OCLC y Forest Press, para evaluar la capacidad de CDD como instrumento de recuperación por materia en el catálogo en línea. Los resultados han sido informados en distintos trabajos (Drabenstott et al., 1990; Markey, 1987). El procedimiento propuesto en este caso es el siguiente:

- 1) el registro documental se copia de un servicio de procesos técnicos, incluyendo la notación de CDD;
- 2) la biblioteca posee una versión de CDD en línea con la notación y los términos verbales equivalentes (denominación de la clase, términos del índice relativo, notas, etcétera);
- 3) el sistema hace coincidir la notación del registro documental con la notación de CDD en línea y los términos verbales equivalentes de esta última se copian automáticamente en el registro bibliográfico.

La cantidad de términos verbales por registro aumentó en forma sustancial con un promedio de 19.6. La denominación de la clase y los términos del índice relativo de CDD han resultado tan adecuados para la recuperación como LCSH y las palabras claves de título. Además, dado que CDD es un esquema jerárquico, se ofrece al usuario la opción de explorar clases más genéricas o más específicas. Mucho antes que las experiencias anteriores—en 1968—el American Institute of Physics de Estados Unidos desarrolló un proyecto de investigación sobre el uso de CDU (edición especializada para ciencia y tecnología nuclear) en sistemas automatizados, incluyendo un catálogo en línea denominado AUDACIOUS.

En este catálogo, el usuario ingresa sus términos de búsqueda en lengua natural y el sistema le muestra en pantalla todas las clases de CDU cuyas denominaciones incluyan dichos términos. Una vez que el usuario selecciona una clase, el sistema exhibe el esquema jerárquico de CDU, permitiendo la exploración de lo general a lo particular y viceversa (Buxton, 1990).

Un desarrollo similar es el de ETHICS de Zurich, Suiza. El sistema consta de un componente jerárquico de conceptos; a cada uno de estos conceptos le corresponde una notación de CDU y numerosos términos verbales, el principal de los

cuales es considerado como un descriptor y el resto como sinónimos, formando entre todos un vocabulario controlado. En adición, los términos verbales se encuentran en tres idiomas (inglés, francés y alemán). Cuando el usuario ingresa sus términos de búsqueda puede optar por la lengua natural o por el vocabulario controlado; luego el sistema le muestra en pantalla todas las clases de CDU cuyas denominaciones incluyan los términos ingresados, permitiéndole acceder a una exhibición sistemática en pantalla y explorar las clases más genéricas o más específicas. El usuario también puede solicitar su búsqueda por la notación, en cuyo caso el sistema exhibe directamente el orden sistemático. ETHICS cuenta con más de 50.000 clases de CDU y 1.3 millones de descriptores (Buxton, 1990).

Durante mucho tiempo se ha postulado que los lenguajes precoordinados (LCSH, LCC, CDD, CDU) no son aptos para la computadora. Sin embargo, la Reunión Satélite de IFLA sobre Indización, realizada en Lisboa, Portugal, en noviembre de 1993, puso de manifiesto que los sistemas precoordinados han sido preferidos para los catálogos en línea de las bibliotecas generales en numerosos países (Holley, 1993).

Existen dos modalidades para tratar estos lenguajes de indización en la computadora. La primera consiste en mantener estrictamente su característica de precoordinados, registrando la cadena en un solo campo y recuperando por la técnica del truncado y/o mediante una recuperación palabra por palabra; en la segunda los encabezamientos y sus subdivisiones, así como las clases y sus números auxiliares, se registran en diferentes subcampos, con lo cual es posible tratarlos con un cierto grado de poscoordinación (Bates, 1988; Buxton, 1990; McIlwaine, 1994).

Por último, la posibilidad de que el usuario “explore los estantes” de la biblioteca desde su propia computadora ha sido sugerida como una forma de complementar el acceso por materia, ya que una vez que se ha elegido uno o varios documentos mediante una búsqueda convencional, se pueden revisar las firmas topográficas y localizar otros documentos que se encuentran adyacentes en el estante porque traen sobre la misma materia (Klugman, 1989).

EL CATÁLOGO DE AUTORIDADES

El control de autoridades es un proceso que consiste en mantener la coherencia entre las formas verbales utilizadas para representar los puntos de acceso al catálogo mediante la verificación y uniformidad de esas formas y mostrar las relaciones entre los nombres (personales, institucionales, geográficos), los títulos y las materias, mediante una estructura de referencias cruzadas (Taylor, 1992).

El instrumento para llevar a cabo este control es el catálogo de autoridades; para construirlo existen reglas aplicables a los nombres y títulos, así como reglas aplicables a las materias. Sin embargo, como indica Taylor (1992) desde hace décadas las bibliotecas han abandonado la práctica de controlar los encabezamientos autorizados, tanto para el catálogo de fichas como para el catálogo en línea. En el mejor de los casos, los bibliotecarios utilizan como sustituto el mismo catálogo público.

Con el desarrollo de los sistemas en línea se ha manifestado la necesidad de contar con un adecuado catálogo de autoridades por los siguientes motivos:

En la etapa de almacenamiento le ofrece al bibliotecario dos ventajas:

- 1.- Indizar en pantalla sobre la base de los términos ya autorizados por la biblioteca.
- 2.- Verificar los términos y realizar cambios globales en algunos o en todos los registros, incluyendo la corrección de errores de tipeado y el reemplazo de términos obsoletos.

A su vez, en la recuperación le ofrece al usuario tres posibilidades:

- 1.- Recuperar por los términos exactos, ya autorizados, asignados a cada documento.
- 2.- Navegar en una búsqueda hasta la conclusión deseada, explorando los términos relacionados y los distintos niveles de jerarquía.
- 3.- Conectar automáticamente los sinónimos con los términos de indización, sin necesidad de que el usuario se dé cuenta de esta conexión.

Las versiones en línea de los lenguajes de indización facilitan la creación de los catálogos de autoridades por materia; incluso es posible considerarlos como sustitutos, ya que ofrecen ventajas similares (Chan, 1990), pero existe en ellos una



El catálogo de autoridades y el lenguaje de indización en línea se complementan durante la indización, ya que el primero facilita la asignación de términos autorizados y la tarea de verificación, mientras que el segundo es necesario para asignar términos nuevos.



importante diferencia: no todos los términos incluidos en un lenguaje de indización son utilizados por una biblioteca y a la inversa, todas las bibliotecas tienen la necesidad de crear términos de indización que no se encuentran en el lenguaje de indización, por ejemplo los identificadores de personas, instituciones o lugares geográficos, así como epígrafes nuevos; es decir, el lenguaje de indización posee términos que nunca se utilizarán en la indización y al mismo tiempo carece de otros que ya se han usado y pueden volver a necesitarse. Por el contrario, el catálogo de autoridades incluye todos aquellos términos autorizados por la biblioteca, que reflejan su colección con exactitud.

En consecuencia, el catálogo de autoridades y el lenguaje de indización en línea se complementan durante la indización, ya que el primero facilita la asignación de términos autorizados y la tarea de verificación, mientras que el segundo es necesario para asignar términos nuevos. En la recuperación, el catálogo de autoridades en línea resulta más útil que el lenguaje de indización, por que le ofrece al usuario una garantía documental del 100%.

En los Estados Unidos, la Library of Congress ha desarrollado un módulo de autoridades para el formato MARC en el que se pueden registrar las formas autorizadas de los encabezamientos y las referencias cruzadas, así como las notas de alcance correspondientes. A su vez, junto con otras bibliotecas cooperantes ha elaborado un catálogo de autoridades para nombres y títulos (LCAF) y otro para materia (LCSAF), disponibles en línea, en microfichas y en CD-ROM.

Aunque el control de autoridades se ha recomendado siempre para los términos verbales, la Library of Congress también ha desarrollado otro módulo del formato MARC para el control de las notaciones autorizadas. El módulo fue diseñado entre 1987 y 1988, en estrecha colaboración con los responsables tanto de LCC como de CDD que son los esquemas de clasificación más difundidos en Estados Unidos y permite registrar la notación, la denominación de la clase, las referencias cruzadas y las notas de alcance (Guenther, 1992).

LOS LENGUAJES DE INDIZACIÓN

LCSH

Durante mucho tiempo los bibliotecarios norteamericanos han expresado sus críticas y propuestas respecto a LCSH (Bloomfield, 1993; Buckland, 1992; Cochran, 1984; Gerhan, 1989; Holley y Killheffer, 1982; Klugman, 1989; Reynolds, 1989; Studwell, 1993) e incluso alguno se ha permitido sugerir la eutanasia para esta lista de encabezamientos (Henige, 1987); las críticas aumentaron significativamente durante el periodo 1970-1979 y más aún en 1980-1989 (Shubert, 1991), coincidiendo con la implementación y evaluación de los catálogos en línea. Entre los defectos detectados se incluyen la desactualización terminológica, la falta de coherencia y especificidad de los encabezamientos, la carencia de principios teóricos para su construcción, así como su redacción de referencias cruzadas.

Otro de los requerimientos que se ha hecho a la Library of Congress —aunque sin éxito hasta el presente— ha sido la redacción de un código para los encabezamientos de materia, en el que coinciden diversos autores como Chan (1989), Reynolds (1989) y Studwell (1993). Sin embargo, desde 1984 se publica un manual para la asignación de epígrafes, que ya cuenta con cuatro ediciones (Pietris, 1987; Taylor, 1992).

En 1986 LCSH cambió la famosa y centenaria estructura de referencias cruzadas inventada por Charles A. Cutter por códigos de tesoro: *see* (véase) por USE, *see also* (véase además) por BT y RT, X por UP, XX por NT y RT. Si bien este cambio ha significado una ayuda por que permite distinguir ciertas relaciones jerárquicas y asociativas, LCSH no ha sido convertida en un tesoro; el estudio a fondo de las relaciones entre los encabezamientos queda aún pendiente y hay que tener en cuenta que los cambios se hicieron en forma mecanizada (Bates, 1988; Dykstra, 1988; Rolland-Thomas, 1993).

LCSH se publica una vez al año en versión impresa, trimestralmente en microfichas o en CD-ROM y por se man a en cinta magnética. También ha sido traducida o adaptada a varios idiomas. En español existe una adaptación que es la Lista de Encabezamientos de Materia para Bi-

bliotecas (LEMB), cuya segunda edición fue publicada en 1985.

LCC

Como es sabido, se trata de un esquema de clasificación universal y jerárquico, pero enumerativo, no sintético. Está dividido en 45 esquemas individuales, algunos de los cuales tienen su propio índice, pero no existe un índice general. El sustituto obvio son los encabezamientos de LCSH, ya que la mayoría de éstos, aunque no todos, remite a las clases de LCC. Distintas experiencias en los Estados Unidos han vinculado ambos lenguajes de indización (Broadbent, 1989; Larson, 1991; Micco, 1991).

Nancy Williamson ha trabajado en las características y componentes estructurales de LCC para preparar la versión en línea (Chan, 1990), tarea más que difícil por la naturaleza enumerativa de este esquema. El módulo del formato MARC para control de autoridades de los datos de clasificación permite que las bibliotecas puedan incorporar estas notaciones, pero Guenther (1991) advierte que una versión legible por computador llegaría a tener cerca de 500 000 clases y presentar serios problemas en cuanto a la capacidad de almacenamiento y velocidad de recuperación.

LCC ha sido traducida al español y es utilizada en diversos países hispanohablantes.

CDD

Desde la aparición de la decimoctava edición de 1979 que aumentó las tablas auxiliares de una a siete —convirtiendo a CDD en una clasificación capaz de ordenar la información y no sólo los estantes— y debido a la extensión que alcanzan las notaciones, más de un tercio de las bibliotecas norteamericanas han re-clasificado sus colecciones con LCC para la signatura topográfica (Chressanthis, 1995; Taylor, 1992). Sin embargo, como ya se mencionó, se han llevado a cabo importantes experiencias que han revalorado a CDD como instrumento de almacenamiento y recuperación de información (Drabenscott, 1990; Markey, 1987). En 1988 el OCLC adquirió la Forest Press y lo gró una versión en línea con fines meramente editoriales. En 1993 se publicó por primera vez *Electronic Dewey*, que es la versión en CD-ROM de CDD para in-

dización y recuperación; está basada en la vigésima edición y cuenta con nueve índices (Trotter, 1995). Por otra parte, el formato MARC incluye tanto la notación de CDD asignada por la Library of Congress como la notación asignada por cada biblioteca individual. Además, el módulo del formato MARC para el control de autoridades de los datos de clasificación permite que cada biblioteca pueda incorporar sus notaciones autorizadas (Guenther, 1991).

La tercera y última edición de CDD en español ha sido publicada en 1995 y se basa en la vigésima edición en inglés.

CDU

En 1993 la FID anunció oficialmente la aparición de la versión de CDU legible por computadora, basada en el Master Reference File, base de datos del Consorcio CDU que cuenta con 60.000 clases. Diversas experiencias internacionales han demostrado la utilidad de CDU para el almacenamiento y recuperación de la información en sistemas computarizados (Buxton, 1990; McIlwaine, 1994; Strachan y Oomes, 1995).

La última edición abreviada española es la sexta de 1991. La Asociación Española de Normalización (AENOR), que es el organismo responsable de CDU en nuestra lengua, ha publicado ya una versión informática de esta sexta edición abreviada y espera publicar la primera edición media en español a fines de 1995.

Otros lenguajes de indización

Los principales avances en relación con los lenguajes de indización incluyen los tesauros, las clasificaciones facetadas, PRECIS y la conciliación de distintos lenguajes.

El tesoro es un lenguaje de indización cuyos términos conforman un vocabulario controlado y se estructuran en un sistema de relaciones equivalentes, jerárquicas y asociativas, de manera que un mismo lenguaje proporciona el acceso alfabético y el acceso sistemático. Nancy Williamson (1989a) ha señalado que los diseñadores de sistemas de información han ignorado las ventajas del componente sistemático del tesoro, que puede ser de gran utilidad para el usuario durante la recuperación, mediante la exhibición de las relaciones jerárquicas y asociativas. Dado su carácter poscoordinado, el

tesauro se adapta mejor al uso de los operadores booleanos en la computadora y la indización resulta menos costosa. Desde su aparición en la década de los cincuenta se ha registrado una proliferación de tesauros especializados, pero desafortunadamente no existe un tesauro universal para las bibliotecas generales.

En cuanto a las clasificaciones facetadas se ha demostrado su eficiencia en sistemas poscoordinados, en la medida en que las clases producidas por el análisis facial estén representadas por una notación tal que permita poscoordinarla con los operadores booleanos, al mismo tiempo que posibilite la exploración jerárquica por medio del truncado (Gödert, 1991). En 1977 se comenzó a publicar en Gran Bretaña la segunda edición de la Bliss Bibliographic Classification (BBC), bajo la dirección de Jack Mills; como es sabido, se trata de un esquema universal y facetado.

La mayoría de los esquemas individuales ya han sido publicados, pero el esquema completo se encuentra aún en curso de publicación (Foskett, 1982; Taylor, 1992; Thomas, 1995). Otra clasificación universal y facetada es la Colon Classification (CC), cuya séptima edición fue publicada en 1988 por la Sarada Ranganathan Endowment; a pesar de sus sesenta años de existencia, este esquema no ha logrado imponerse en las bibliotecas generales ni competir en difusión y uso con las tradicionales CDD y CDU. Ni BBC ni CC cuentan con versiones legibles por computadora, ni han sido traducidas al español.

Por su parte PRECIS es un sistema creado en Gran Bretaña por Derek Austin, que entró en servicio en la British National Bibliography en 1971. Más que un lenguaje de indización se trata de un procedimiento basado en el concepto de vocabulario *open-ended*, es decir que los términos pueden ser admitidos en cualquier momento en que aparezcan en la literatura.

Una vez que el documento ha sido analizado, los conceptos son ordenados como un título en lengua natural, formando una cadena de términos, cada uno de los cuales está precedido por un operador terminado. Aquellos términos que aparecen en la posición líder son marcados y las estructuras semánticas se añaden en forma de referencias de *véase además* para

cualquiera de los términos de la cadena. También se incluyen los códigos de manipulación para la computadora y ésta se encarga del ordenamiento (Dykstra, 1989; Lancaster, 1991; Rowley, 1988). Lamentablemente, PRECIS resulta costoso de bido al tiempo que con su método de análisis sintáctico y la codificación de las cadenas; a fines de 1990 la British Library se vio forzada a sustituirlo por un sistema menos sofisticado denominado COMPASS (Holley, 1993; Taylor, 1992); por el mismo motivo la Library of Congress nunca adoptó PRECIS, a pesar de ciertas presiones de la comunidad bibliotecaria norteamericana (Holley y Killheffer, 1982).

Conciliación de lenguajes de indización

Con la proliferación de tesauros y clasificaciones especializadas, se ha producido un aumento en la incompatibilidad de los catálogos en línea, tanto en el nivel de especificidad, como en la exhaustividad, los términos compuestos, los sinónimos y las relaciones entre los términos. La compatibilización o conciliación de estos lenguajes incluye las técnicas de conexión y fusión, así como los macrotesauros (Aitchison y Gilchrist, 1987; Niehoff y Mack, 1985; Rada, 1987; Taylor, 1992).

Ejemplos de esta conciliación son el sistema de conexión de vocabulario desarrollado por Battelle Columbus Laboratories de Estados Unidos (Niehoff y Mack, 1985); el sistema unificado de lenguajes médicos de la U.S. National Library of Medicine (Squires, 1993), o el sistema operacional de la NASA que conecta sus propios términos de indización con aquellos del U.S. Department of Defense, mediante una tabla de traducción léxico y un programa que determina las reglas a seguir (Silvester y Klingbiel, 1993).

Douglas J. Foskett (1991) ha manifestado que BBC puede ser el mejor lenguaje de conciliación. Este esquema facetado tiene una doble función: a) dado que sus clases principales han sido desarrolladas en esquemas individuales con un alto nivel de especificidad, es posible indizar en profundidad aún la información más específica, y b) todos los esquemas individuales forman un solo esquema universal, totalmente compatible, porque al



Con la proliferación de tesauros y clasificaciones especializadas, se ha producido un aumento en la incompatibilidad de los catálogos en línea



habersidoelaboradoconlamismametodología de análisis facetal, estructurado con las mismas reglas y dotado de un mismo sistema de notación, garantiza una homogeneidad global.

CONCLUSIONES

ComoseñalaMandel(1991),lanaturaleza inexacta y subjetiva de la búsqueda por materia requiere múltiples rutas de acceso para facilitar la coincidencia entre los términos del usuario y el vocabulario del sistema. Por su parte, Chan (1990) y Larson (1991b) están de acuerdo con que no basta incluir más términos, sino en combinar distintos tipos de términos ya sean palabras claves de la lengua natural, así como términos controlados de acceso verbal (encabezamientos de materia o descriptores) y de acceso sistemático (notaciones).

Seresumen a continuación las diferentes propuestas para mejorar la indización por materia de las monografías en los catálogos en línea de bibliotecas generales (Larson, 1991a):

- 1.- Mayor cantidad de términos verbales controlados (encabezamientos de materia o descriptores): asignar más términos verbales controlados por registro y explotar las versiones en línea de los catálogos de autoridades o los lenguajes de indización para proporcionar mayor coherencia a los indizadores y orientación a los usuarios.
- 2.- Mayor cantidad de palabras claves (lengua natural): explotar los términos individuales de cada campo por medio de palabras claves, añadir palabras claves de título y/o resumen en

cada registro y, en el caso de las monografías, incluir palabras claves de sumarios e índices analíticos.

- 3.- Mayor cantidad de notaciones: asignar notaciones a todos los registros y añadir los términos verbales derivados de los esquemas de clasificación y sus índices.

Por otro lado, si bien los lenguajes de indización tradicionales (LCSH, LCC, CDD, CDU) han sido muy criticados por su carácter precoordinado, su dificultad para procesarlos en la computadora, la obsolescencia de su terminología, su ordenamiento por disciplinas y no por conceptos, etcétera, cuentan con una amplia difusión internacional, han sido traducidos a varios idiomas y administrados durante décadas por organizaciones responsables (Library of Congress, Forest Press, FID), que han garantizado su continuidad.

Las dificultades económicas de este último cuarto de siglo, que en mayor o menor grado afectan a las bibliotecas de todo el mundo, impiden o hacen muy difícil que puedan crearse nuevos lenguajes de indización universales más eficientes o modificar sustancialmente los tradicionales y, más aún, poner en práctica una posible re clasificación o reindización de los miles de millones de registros bibliográficos existentes (Chan, 1990; Holley, 1989; Mandel, 1991).

De manera que por el momento no se vislumbran cambios en el uso de los lenguajes de indización para los catálogos en línea de las bibliotecas generales, sino un uso más eficiente de los actuales mediante el aumento de los puntos de acceso por materia en cada registro y su comple-

mento con la lengua natural (Williamson, 1989b).

En Argentina —y tal vez suceda lo mismo en otros países latinoamericanos— la mayoría de los catálogos informatizados de las bibliotecas generales no cuenta aún con un acceso remoto que facilite la consulta al usuario; en consecuencia, no se puede conocer con exactitud el desempeño de los usuarios durante la búsqueda por materia.

Sin embargo, los resultados preliminares de una evaluación de estos sistemas en nuestro medio (Pichini y Martínez, 1995) demuestran que los puntos de acceso por materia son prácticamente los mismos que en la ficha: uno o dos encabezamientos y un número clasificatorio; excepcionalmente palabras claves de título y en ningún caso el resumen. De esta forma, es posible anticipar un alto índice de fracasos en la búsqueda por materia, tal como se ha podido detectar en otros países (Larson, 1991b; Su, 1994).

Las propuestas del formato aumentado que se han mencionado anteriormente, en especial la inclusión de un resumen, implican un aumento en los costos, tanto en la capacidad de almacenamiento y procesamiento de datos, como en recursos humanos calificados, que quizás muchas bibliotecas no estén en condiciones de afrontar. Por otra parte, los lenguajes documentales universales en español (LEMB, LCC, CDD, CDU) demoran mucho tiempo en actualizarse y, hasta donde se sabe, sólo CDU está disponible en versión informática en nuestro idioma, con lo cual las dificultades para la indización son aún mayores.

Quizás las experiencias obtenidas en otros países, tal como se ha intentado presentar en esta revisión, puedan contribuir a una amplia discusión de estos problemas, a la prevención de los mismos errores y al diseño de catálogos en línea más eficientes para los usuarios de nuestras bibliotecas.

BIBLIOGRAFÍA

- Aitchison J, Gilchrist A. *Thesaurus construction*. 2 ed. London: Aslib, 1987.
- Bates M. "How to use controlled vocabularies more effectively in online searching". *Online* 1988; 12(6):45-56.
- Bates M. "Rethinking subject cataloging in the online environment". *Library references and technical services* 1989; 33(4):400-12.

- Bloomfield M. "A look at subject headings: a plea for standardization". *Cataloging and classification quarterly* 1993; 16(1):119-24.
- Broadbent E. "The online catalog: dictionary, classified, or both?" *Cataloging and classification quarterly* 1989; 10(2):105-17.
- Buckland MK. "Agenda for online catalog designers." *Information technology and libraries* 1992; 11(2):157-62.
- Buxton AB. "Computer searching of UDC numbers." *Journal of documentation* 1990; 46(3):193-127.
- Carlyle A. "Matching LCSH and user vocabulary in the library catalog." *Cataloging and classification quarterly* 1989; 10(2):37-63.
- Chan LM. "A subject cataloging code?" *Cataloging and classification quarterly* 1989; 10(2):199-202.
- . "Subject analysis tools online: the challenge ahead." *Information technology and libraries* 1990; 9(3):258-62.
- Chresanthis JD. "The reclassification decision: Dewey or Library of Congress." *Cataloging and classification quarterly* 1995; 19(3/4):169-82.
- Cochrane PA. "Modern subject access in the online age." *American libraries* 1984; 15(2):80-3; 15(3):145-8, 150; 15(4):250-2, 254-5; 15(5):336-9; 15(6):438-41; 15(7):527-9.
- Cousins SA. "Enhancing subject access to OPACS: controlled vocabulary vs. natural language." *Journal of documentation* 1992; 48(3):291-309.
- Culkin PB. "Rethinking OPACS: the design of assertive information systems." *Information technology and libraries* 1989; 8(2):172-7.
- Drabenstott KM, y otros. "Analysis of a bibliographic database enhanced with a library classification." *Library resources and technical services* 1990; 34(2):179-98.
- Dubois CPR. "Free text vs. controlled vocabulary: a reassessment." *Online review* 1987; 11(4):131-6.
- Dykstra M. "LC subject headings disguised as a thesaurus." *Library journal* 1988; 113:42-6.
- . "PRECIS in the online catalog." *Cataloging and classification quarterly* 1989; 10(2): 81-94.
- Foskett D. "Concerning general and special classification." *International classification* 1991; 18(2):87-91.
- Foskett AC. *The subject approach to information*. 4 ed. London: Clive Bingley, 1982.
- Gerhan DR. "LCSH in vivo: subject searching performance and strategy in the OPAC era." *Journal of academic librarianship* 1989; 15(2):83-9.
- Gödert W. "Facet classification in online retrieval." *International classification* 1991; 18(2):98-109.
- Guenther RS. "The development and implementation of the USMARC format for classification data." *Information technology and libraries* 1992; 11(2):120-31.
- Henige D. "Library of Congress subject headings: is euthanasia the answer?" *Cataloging and classification quarterly* 1987; 8(1):7-19 y 21-3.
- Holley RP, Killheffer RE. "Is there an answer to the subject access crisis?" *Cataloging and classification quarterly* 1982; 1(2/3):125-33.
- . "Subject access in the online catalog." *Cataloging and classification quarterly* 1989; 10(2):3-8.
- Holley RP. "Report on the IFLA Satellite Meeting Subject Indexing: Principles and Practices in the 90's, August 17-18, 1993, Lisbon, Portugal." *Cataloging and classification quarterly* 1993; 18(2):87-100.
- Kern-Simirenko C. "OPAC user logs: implications for bibliographic instruction." *Library hi tech* 1983; 1(3):27-35.
- Klugman S. "Failures in subject retrieval." *Cataloging and classification quarterly* 1989; 10(2):9-35.
- Knutson G. "Subject enhancement: report on an experiment." *College and research libraries* 1991; 52(1):65-9.
- Lancaster FW. *Indexing and abstracting*. Champaign: University of Illinois, 1991.

- Larson RR. "Classification clustering, probabilistic information retrieval and the online catalog." *Library quarterly* 1991a; 61(2):133-73.
- Larson RR. "The decline of subject searching: long-term trends and patterns of index use in an online catalog." *Journal of the American Society for Information Science* 1991b; 42(4):197-215.
- McIlwaine IC, Buxton A. *Guía para uso de la CDU: Clasificación decimal universal*. Madrid: AENOR, 1994.
- Mandel CA. "Cataloging for access." *Reference librarian* 1991; (34):61-8.
- Markey K. "Searching and browsing the Dewey Decimal Classification in an online catalog." *Cataloging and classification quarterly* 1987; 7(3):37-68.
- Micco M. "The next generation of online public access catalogs: a new look at subject access using hypermedia." *Cataloging and classification quarterly* 1991; 13(4):103-29.
- Niehoff R, Mack G. "The vocabulary switching system: description of evaluation studies." *International classification* 1985; 12(1):2-6.
- Peters TA. "When smart people fail: an analysis of the transaction log of an online public access catalog." *Journal of academic librarianship* 1989; 15(5):267-73.
- Pichini M, Martínez AM. *Catálogos en línea de bibliotecas bonaerenses: resultados preliminares de una encuesta* [Resumen]. Curso de Actualización sobre Acceso en Línea a los Catálogos de Bibliotecas. La Plata: UNLP, 1995.
- Pietris MKD. "LC: new manual, not code, needed." *American libraries* 1987; 18:958.
- Rada R. "Connecting and evaluating thesauri: issues and cases." *International classification* 1987; 14(2):63-9.
- Reynolds SJ. "In theory there is no solution: the impediments to a subject cataloging code." *Library quarterly* 1989; 59(3):223-38.
- Rolland-Thomas P. "Thesaural codes: an appraisal of their use in the Library of Congress subject headings." *Cataloging and classification quarterly* 1993; 16(2):71-91.
- Rowley JE. *Abstracting and indexing*. 2 ed. London: Clive Bingley, 1988.
- Settel B, Cochrane PA. "Augmenting subject descriptions for books in online catalogs." *Database* 1982; 5(4):29-37.
- Shubert SB. "Critical views of LCSH ten years later: a bibliographic essay." *Cataloging and classification quarterly*, 1992; 15:37-97.
- Silvester JP, Klingbiel PH. "An operational system for subject switching between controlled vocabularies." *Information processing and management* 1993; 29(1):47-59.
- Squires SJ. "Access to biomedical information: the unified medical language system." *Library trends* 1993; 42(1):127-51.
- Strachan PD, Oomes FMH. "Universal decimal classification update." *Cataloging and classification quarterly* 1995; 19(3/4):119-32.
- Studwell WE. "LC's head in the sand, or, why the subject cataloging manual is not enough." *Technical services quarterly* 1993; 10(3):45-50.
- Su SF. "Dialogue with an OPAC: how visionary was Swan son in 1964?" *Library quarterly* 1994; 64(2):130-61.
- Svenonius E. "Unanswered questions in the design of controlled vocabularies." *Journal of the American Society of Information Science* 1986; 37(5):331-40.
- Taylor A, Bohdan S. *Wynar: Introduction to cataloging and classification*. 8 ed. Englewood: Libraries Unlimited, 1992.
- Thomas AR. "Bliss classification update." *Cataloging and classification quarterly* 1995; 19(3/4):105-18.
- Trotter R. "Electronic Dewey: the CD-ROM version of the Dewey Decimal Classification." *Cataloging and classification quarterly* 1995; 19(3/4):213-34.
- Williamson NJ. "The role of classification in online systems." *Cataloging and classification quarterly* 1989a; 10(1):95-104.
- . "Classification in the computer age." *International classification* 1989b; 16(1):30-4.