

La automatización de la CDU y los sistemas de gestión de bases de datos: El análisis de la estructura de la información

Xavier Agenjo
Pilar Benedito
Dolores del Castillo
María Jaudenes

Se describe la estructura de la base de datos ARIADNA, de la Biblioteca Nacional de Madrid, la cual permite acceder a la información por autor, título, materia, etc., aprovechando las ventajas que ofrece la CDU para la organización y gestión de la colección, como son las relaciones jerárquicas, las asociativas y las de equivalencia. De estas últimas, las de sinonimia con términos equivalentes de otros lenguajes de indización (como los encabezamientos de materia) son fundamentales para que el usuario pueda llegar a una notación de CDU. Se apuntan las limitaciones de las actuales generaciones de computadoras, para asimilar la CDU con todas sus posibilidades de explotación. Se señala que el uso de sistemas de clasificación en catálogos *on line* está todavía en fase de experimentación.

The structure of ARIADNA database, at the National Library in Madrid, is described. This database allows the information access by author, title, subject, etc.; taking advantages of the UDC possibilities for the organization and management of libraries, through its hierarchical associative, and equivalence relations. Among the latter, the synonymy with equivalent terms from the other indexing languages (ie subjects headings) is an option in order to have access to any UDC notation. The limitations of the present generations of computers to afford the maximal exploitation of the UDC are pointed out, although the use of classification systems for on-line catalogs is realized to be on an experimental basis.

Introducción

Hace unas decenas de años, cuando un lector de la Biblioteca Nacional deseaba consultar un libro debía indicar al funcionario que se encontraba enfrente de él, al otro lado del mostrador, el nombre del autor, el título de la obra y cuantos datos pudieran identificarla para que esta llegara a sus manos, tras la correspondiente búsqueda en el viejo índice de la Biblioteca Nacional cuyos centenares de miles de fichas se apretaban en unos mesetones a los que solo tenían acceso los bibliotecarios.

Años más tarde, se tomó la verdaderamente revolucionaria medida, por más que a nosotros nos resulte inconcebible que no fuera así, de poner al público numerosos ficheros, que contenían fichas duplicadas del viejo índice, ordenadas unas por autores, títulos y materias en un catálogo diccionario y otras por CDU en un catálogo sistemático.

En los años 60 se comenzó a hablar de la mecanización de los catálogos de la biblioteca; en los años 70 se trazaron los primeros proyectos; en los 80 se realizaron las primeras tentativas y al inicio de la década de los años 90 (1) se podrá culminar la mecanización del proceso catalográfico.

Como es lógico, este proceso de creación de registros, o dicho más técnicamente, de la actualización de la base de datos, requiere la posibilidad de buscar la información necesaria para **cumplimentar una ficha**. Y gracias a la compleja estructura de la base de datos diseñada según **el método de entidad/relación**, que permite el procedimiento conocido en la moderna teoría de la **comunicación** como navegar por la información, es posible llegar al dato deseado. Por desgracia, la **contradicción entre potencia y flexibilidad** en la recuperación de la información de una base de datos y **facilidad en su manejo** es muy difícil de salvar y es necesario interponer entre el lector

no experimentado en el manejo de complejos sistemas informáticos y la enorme cantidad de datos finamente estructurados entre sí, un lenguaje, una interfaz que de forma amigosa e inteligible a primera vista haga viable la consecución inmediata por parte del lector de la información que necesita. Esta interfaz, conocida técnicamente con el nombre de OPAC, siglas en inglés de *on line public access catalogue*, es el procedimiento por el cual el lector conoce por dentro a ARIADNA, nombre de la base de datos de la Biblioteca Nacional.

En la actualidad, ARIADNA permite acceder a la información por autor, títulos, materias, etc., y, además proporciona los medios necesarios para conseguir más información de los registros bibliográficos mediante un tratamiento adecuado del sistema de clasificación utilizado, la CDU. Esto, unido a las posibilidades que ofrece en sí misma la estructura de la CDU y el hecho de que sea el sistema de clasificación más usado no solo en la Biblioteca Nacional sino también en España, nos animó a realizar este estudio.

Antecedentes históricos

Prácticamente desde el inicio de la era de los ordenadores, los bibliotecarios advirtieron las posibilidades que podían tener para el tratamiento de la CDU. Sin embargo, los problemas, estrictamente computacionales, de la primera y segunda generación, dificultaron el tratamiento de una estructura tan compleja como la de la CDU. En esta primera fase, la información solo ayudó a la Clasificación Decimal Universal en los trabajos de mantenimiento y, sin duda alguna, facilitó de forma considerable la preparación de índices. Más difícil fue el planteamiento del control de vocabulario y el establecimiento de un lenguaje de recuperación propio a la creación de tesauros aprovechando la propia estructura de la CDU.

Donde sí se mostró eficaz la CDU fue en el proceso de difusión selectiva de la información. Efectivamente, era muy fácil utilizar cualquier notación de la Clasificación como una ecuación de búsqueda que resumiese el perfil documental de un usuario. Incidentalmente puede comentarse que las primeras transferencias masivas que la Biblioteca Nacional ha llevado a cabo de su base de datos ARIADNA han sido justamente a través de una notación decimal.

Tampoco la recuperación en línea resultó tan potente como la estructura de la CDU y la información en ella contenida prometían. No es fácil incluir la notación CDU bajo el criterio de alfanumérico o textual ni, desde luego, bajo el de clave puramente numérica, aunque así lo sea en la mayoría de los casos. Sobre gran parte de estos problemas nos extenderemos más adelante.

Las primeras dificultades con las que se topó el tratamiento automatizado de la CDU estaban basadas en las longitudes variables, lo que dificultaba su definición como campo clave, así como el tratamiento de notaciones compuestas. De forma anecdótica puede citarse la falta de signos característicos de la CDU en las primeras impresoras disponibles en el mercado.

A pesar de que la evolución de los ordenadores ha resuelto mucho de los problemas iniciales, quedan otros pendientes como es el hecho de carecer de un formato normalizado para los datos correspondientes a la CDU, aunque existen el USMARC para la Library of Congress Classification y para la Dewey Decimal Classification y está en estudio un formato internacional para clasificación por parte de la Sección de Indización y Clasificación de la Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas (IFLA).

Modelo

A la hora de realizar un trabajo de la envergadura del que se propone en este artículo es muy importante escoger el procedimiento a seguir previamente a llevar a cabo el trabajo propiamente dicho.

Para ello el método de entidad/relación, bien conocido entre los analistas e informáticos puede ser sumamente útil. Lo primero que se va a llevar a cabo es una modelización conceptual que comporte una definición formal del modelo de datos. Este modelo de datos consiste en un conjunto de

conceptos, reglas y convenciones que nos permiten describir los datos de una parcela del mundo real. Es obvio que la CDU comporta una estructura de datos previa.

El modelo de datos tiene que enfocarse desde dos visiones distintas. La primera de ellas hace referencia al conjunto de los objetos que forman parte del modelo, es decir, a entidades con sus atributos; en segundo lugar, al conjunto de asociaciones entre ellos, es decir, sus interrelaciones; y, por último, al conjunto de restricciones inherentes al propio modelo, o bien opcionales, basados fundamentalmente en el rendimiento del sistema.

La segunda de las perspectivas citadas es la dinámica, es decir, el conjunto de operadores que permiten la recuperación y la actualización de los datos. En lo que se refiere a los operadores para la recuperación de los datos, el sistema ARIADNA cuenta ya con un conjunto de ellos como podrían ser TE, TG, TR, USE y UP, a los que se podrían añadir todos aquellos que fueran necesarios.

Los objetos de la estructura de metamodelo fundamentalmente pueden dividirse en cuatro: las entidades, por ejemplo, una notación de la CDU, los atributos, que son las propiedades de las entidades, por ejemplo, la obligatoriedad o no de su cumplimentación, los dominios, sobre los cuales se definen los atributos que, en nuestro caso serían claramente las distintas tablas auxiliares de la CDU y, por último, las interrelaciones formadas por las vinculaciones o asociaciones entre los distintos objetos.

La definición de la entidad CDU en la metaestructura de ARIADNA

El subsistema de definición de metaestructura de ARIADNA permite la creación y mantenimiento de tantas entidades, llamadas en ARIADNA entes, como sea necesario para la adecuada organización y gestión de la colección y de la información que se maneja en la Biblioteca Nacional. Se puede definir un ente como aquel objeto del mundo real del cual se quiere guardar información y que posee unas características o valores concretos, atributos, que lo identifican y lo diferencian de los demás.

Los sistemas de clasificación (SC en adelante) presentan características estructurales diferentes en lo que se refiere a tipo de vocabulario empleado (numérico o alfanumérico), más o menos enumerativos, sintaxis para la expresión de nuevos conceptos, signos de puntuación o criterios de ordenación de las notaciones.

Además, dentro del registro bibliográfico de ARIADNA, las notaciones de los distintos SC, el Inconclass, por ejemplo, se codifican en elementos diferentes, por lo que sería necesario controlar que las notaciones de cada uno de ellos se asignan al elemento correcto del registro bibliográfico.

Todo ello determina la existencia de un ente distinto para cada uno de los SC que se desee controlar y vincular mediante asociaciones con notaciones de otros sistemas, términos de otros lenguajes de indización y otros tipo de registros, como los bibliográficos.

La existencia del ente CDU está motivada por la necesidad de un diseño adecuado que permita la introducción, mantenimiento, búsqueda y recuperación de los distintos elementos que componen una notación CDU, aprovechando al máximo las ventajas que ofrece la estructura de este sistema de clasificación: notación básica, auxiliares especiales, auxiliares comunes, signos que determinan la clase de relación entre dos notaciones, etc.

La información que constituye un ente está descompuesta en elementos o unidades mínimas de información. Cada uno de ellos tiene un nombre que lo identifica y unas características asociadas. Sus características intrínsecas serán las de la información que representen.

Los elementos se agrupan entre sí formando grupos homogéneos de información. Estos grupos están determinados por la normativa que rige la catalogación y la redacción de registros, así como por las necesidades de proceso de la biblioteca. Estos grupos son homólogos a los campos MARC. Los elementos que pueden subdividirse en otros, por ejemplo, el grupo de informaciones que componen una notación compleja, se denominan elementos no terminales. Los elementos que no pueden descomponerse en otros, por ejemplo, el que contiene el auxiliar de lengua dentro de la notación, se llaman elementos terminales y son los verdaderos portadores de la información, homólogos a los subcampos MARC.

El análisis de los elementos que constituyen el ente CDU se puede considerar desde dos puntos de vista:

1) El comportamiento dentro del ente en función de:

- Jerarquía. Por la cual unos elementos se subordinan a otros, formando unos grupos homogéneos.
- Obligatoriedad. Por la cual un elementos puede quedar o no vacío según se defina como opcional u obligatorio
- Repetibilidad. Determina que un elemento pueda utilizarse una o más veces dentro de su nivel jerárquico superior.

2) Las características, o atributos, que posee la información contenida en los elementos terminales. Los atributos que definen los elementos son:

- Tipo de información: textual, numérica, códigos alfanuméricos, punteros.
- Longitud (para aquellas informaciones que tienen una longitud fija o máxima).
- Tratamiento que va a recibir la información:
 - No se va a establecer ningún tipo de control sobre la información. Textual.
 - Es un valor asignado cuando se crea el registro por ejemplo, la fecha de entrada en el sistema.
 - Es una información formateada, con una longitud y un contenido controlado, bien por valores representados en una tabla, bien por la especificación de un rango de valores máximos/mínimos entre los que debe estar la información del elemento, bien mediante la realización de operaciones *ad hoc* utilizando unas variables dadas.
 - Es un puntero, es decir, un valor que identifica otro registro desde el cual se transfiere información.

Estos atributos intrínsecos de la información que contienen los elementos terminales aparecen definidos en la máscara del elemento. Un elemento terminal puede tener varias máscaras, lo cual permite variar la estructura y las validaciones del elemento en función de la máscara utilizada.

En el ente CDU, como en el resto de los entes que existen en ARIADNA, hay una serie de elementos comunes:

- Número de control (001 IBERMARC). Obligatorio, no repetible y generado por el sistema.
- Número anterior (010, 035 IBERMARC). Opcional, repetible y textual (sin validación por parte del sistema)
- Fecha de última modificación (005 IBERMARC), con mención de usuario y centro modificador. No repetible y generado por el sistema.
- Fecha de alta en la aplicación (no IBERMARC), con mención de usuario y centro creador. Obligatorio, no repetible y generado por el sistema con la estructura, idéntica al elemento anterior, AAAAMMDDHHMM USUARIO CENTRO.
- Información codificada sobre el propio registro, que corresponde con ciertas posiciones del LEADER en la estructura IBERMARC. Es obligatorio, no repetible y tiene una longitud fija, cada una de cuyas posiciones valida sus valores contra una tabla.
- Información codificada sobre el contenido del registro (008 IBERMARC). Obligatorio, no repetible y con una longitud fija, cada una de cuyas posiciones valida sus valores contra una tabla.

- Información sobre los centros que han intervenido en la creación o modificación del registro y sobre la lengua de catalogación (040 IBERMARC). Obligatorio, no repetible y validado contra tablas (centros y lenguas).
- Código de verificación (042 IBERMARC). Opcional y repetible. Una parte de su contenido la genera el sistema (fecha y hora de verificación), otra parte se valida contra tablas (usuario y centro).

Otro grupo de elementos que está presente en todos los entes de ARIADNA, con diferentes denominaciones, son las notas. A veces son instrucciones para el catalogador sobre cómo aplicar los diferentes auxiliares o construir las notaciones (683 IBERMARC), consejos sobre cómo clasificar determinadas materias, prácticas de clasificación (684 IBERMARC), etc. Otras veces son notas explicativas, que proporcionan información sobre el tipo de materias que se clasificarán bajo esa notación o la distinguen de otras que aparecen similares (680 IBERMARC). Estos campos de notas son opcionales y repetibles, la información que contienen es textual.

Existen también otros elementos, presentes también en otros entes con distintas denominaciones, que sirven para relacionar el registro que se está catalogando con otros ya existentes en la base de datos. Más adelante se tratarán las asociaciones que se pueden establecer entre los registros de CDU y el modo en que se establecen.

Por último, están los elementos que contienen la información que es verdaderamente distintiva del ente. En el ente CDU es el grupo de elementos que contiene la notación.

Para determinar en cuántos elementos se debería descomponer la notación es necesario hacer una breve análisis de los elementos de que puede estar compuesta:

- una o más notaciones básicas vinculadas entre sí por varios signos (':', '+', '/');
- notaciones básicas clasificadas por auxiliares especiales (marcados por '.0' o por '-');
- notaciones básicas calificadas por uno o más auxiliares comunes o por A/Z.

A esta compleja estructura se debe sumar:

- La práctica de separar mediante un punto en grupos de tres los números que constituyen la notación. Este punto se desplaza a la derecha cuando se adicionan números por la izquierda, por ejemplo: (460) es España; precedido de 91, para significar la geografía de España, quedaría 914.60. El punto, sin embargo, no se desplaza cuando lo que se adiciona por la derecha son auxiliares marcados por punto, por ejemplo: 840.9 es literatura catalana; con el auxiliar .09 se expresaría la crítica literaria y quedaría 840.9.09.
- La utilización de auxiliares especiales, restringida a la clase de la que forman parte, que presentan una misma forma, pero con un significado diferente según la clase a la que califiquen, por ejemplo, 54-31 son óxidos; 860-31 es novela española.
- El hecho de que la CDU posea un recurso de ahorro consistente en utilizar un mismo elemento como base para construir notaciones diferentes. Por ejemplo, los auxiliares de lugar, sin paréntesis, sirven para expresar la historia, precedidos de un '9', o la geografía, precedidos de un '91', de ese lugar; a los números que se subdividen como otros. Por ejemplo, el 632.7, insectos nocivos para la agricultura, se especifica subdividiendo como 595.7, entomología.
- El orden de cita establecido para los elementos que componen una notación puede verse alterado.
- La ordenación de las distintas notaciones entre sí, que pueden ser incompatible con los pesos que se hayan establecido para el juego de caracteres con vistas a la ordenación alfabética.

El encadenamiento de los distintos elementos que pueden componer una notación muy compleja proporciona cuatro niveles dentro del ente a los que se denomina subentes:

1) *Notación básica con o sin auxiliares especiales:*

Notación básica. Formaría un grupo con los auxiliares especiales, aún siendo cada uno un elemento terminal independiente. De esta forma se podría validar que con cada una de las clases en que se divide el CDU se pudieran utilizar solo aquellos auxiliares especiales permitidos por el SC, y solo en el modo en que el SC prescribe. Los auxiliares especiales de cada clase estarían validados contra una tabla. Por ejemplo, 016 (bibliografías especiales); 806.0-311 (toponimia en lengua española).

2) *Notación básica con auxiliares comunes o A/Z:*

Elemento-puntero cuyo contenido es el localizador del registro de la notación básica. Un elemento por cada uno de los tipos de auxiliar común. Un elemento por auxiliar facilita la validación del contenido, puesto que cada clase de auxiliar está en una tabla diferente. Además, evita que el indizador tenga que teclear la puntuación asociada. La existencia de una tabla diferente por auxiliar también facilita la utilización del mismo para construir otras notaciones.

El único auxiliar común que no se validaría contra tabla sería el de tiempo, por las múltiples posibilidades de construcción que presenta. Tampoco se validan los valores A/Z; por ejemplo, 016(460)"18".

3) *Notación básica vinculada a otra notación:*

Elemento puntero para el localizador de la notación principal, con o sin auxiliares comunes y A/Z. Este elemento podría ser el mismo para todos los tipos de relación existentes ('+' '/' /'), pero sería más conveniente que hubiera un elemento distinto para cada uno de los tipos de vinculación. De este modo, la puntuación se asignaría por la aplicación y se facilitaría la navegación diferenciada entre ellos. Por ejemplo, 016:5; 5/6.

En este subente 3 habría, además, los elementos necesarios para formar la notación básica con la que se quiere vincular la del elemento puntero. Estos elementos son los mismos que los citados en los párrafos anteriores.

4) *Dos o más notaciones vinculadas, acompañadas de auxiliares comunes o A/Z:*

Elemento-puntero que contiene dos o más notaciones básicas relacionadas. Un elemento para cada auxiliar común y para A/Z. Por ejemplo, 5/6(063).

Relaciones

Los diferentes tipos de relaciones que pueden establecerse entre los registros se denominan asociaciones. El establecimiento de asociaciones entre registros permite la recuperación de unos a partir de otros mediante la función navegación.

En ARIADNA se pueden establecer distintas clases de asociaciones entre los entes existentes en la base de datos, lo que permite su recuperación.

- Asociaciones implícitas. Parte de la información de un registro (origen) pasa a formar parte de otro (destino); por ejemplo, cuando se transfiere un encabezamiento de materia a un registro bibliográfico. La transferencia se realiza a través de un elemento-puntero.
- Asociaciones explícitas. El catalogador da de alta un registro de asociación que consta como mínimo de dos elementos-puntero, el localizador de cada uno de los dos entes (uno origen y otro destino) entre los que se quiere establecer la asociación.

En la definición de asociaciones hay cuatro fases:

- 1) Determinar qué clase de relaciones interesa establecer.
- 2) Determinar cuáles van a ser los dos sentidos de la asociación (la navegación de origen a destino y viceversa), para saber las características que deben reunir el ente-origen y el ente-destino.
- 3) Determinar qué entes van a poder ser origen o destino.
- 4) Establecer los códigos de navegación, comandos que permitirán la recuperación de los registros asociados y que indican el tipo concreto de asociación y el sentido de la misma.

Dejando de lado la distinción entre asociaciones explícitas e implícitas, el análisis de las asociaciones del ente CDU se va a centrar en un intento de categorización y enumeración de las relaciones posibles.

Relaciones jerárquicas

Son las que se establecen entre los distintos subentes de una notación y serían asociaciones claramente implícitas. En ellas, el término genérico sería el de nivel superior, por ejemplo, la notación básica, y el término específico sería el de nivel inferior, por ejemplo, la notación básica acompañada de auxiliares comunes.

Serían jerárquicas también las establecidas con encabezamientos de materia (como lenguaje de indización más empleado en la Biblioteca Nacional) que pertenezcan al campo temático expresado por esa notación. En este sentido la CDU actuaría como un *top term* bajo el que se agruparía todo un grupo de encabezamientos-descriptores organizados semánticamente mediante relaciones jerárquicas y asociativas. Como es lógico, las notaciones que intervendrían en este tipo de asociaciones serían genéricas.

Relaciones asociativas

Con otras notaciones de la CDU que representen puntos de vista diferentes del mismo tema. Por ejemplo, 639.2 (pesca) con 343.772 (delitos de pesca) y con 799.1 (pesca deportiva). Se podrían considerar también relaciones asociativas las que se establecen con los registros bibliográficos y con los registros de fondos y localizaciones en las que se haya utilizado una determinada relación.

Relaciones de equivalencia

Relaciones entre formas alternativas, por ejemplo, 629.1 — 629.3, o bien, relaciones de sinonimia con términos equivalentes de otros lenguajes de indización. Según interese, esta relación se establece en un nivel genérico o bien en cualquier nivel. Este tipo de relaciones permite la creación de una especie de diccionario entre los diferentes lenguajes de indización empleados, que permitirá "traducir" la forma de expresar un concepto en los diferentes sistemas, y de ahí navegar a otros términos relacionados.

Las relaciones de sinonimia establecidas con los encabezamientos de materia son fundamentales para que el usuario pueda llegar a una notación, que está expresada en un lenguaje que le es ajeno, partiendo de un término tomado del lenguaje natural, más o menos controlado, que es el que conoce.

La recuperación de la información en ARIADNA

El módulo de meta estructura de ARIADNA permite generar cuantos puntos de acceso al contenido de la información sean necesarios, y además hacerlo de una forma dinámica, de tal modo que pueda adaptarse a necesidades cambiantes.

En ARIADNA se accede a la información mediante vistas, que permiten la recuperación de la información de uno o más de un ente. Existen, básicamente, dos clases de vistas:

- Alfabéticas. Presentan una línea con los datos de cada registro. Los registros aparecen ordenados entre sí por un criterio alfabético, tal y como se encontrarían en un fichero manual; se busca, también como en un fichero manual, por un valor de inicio.
- De recuperación por palabras. Permite recuperar términos extraídos de un índice que se genera al invertir las palabras contenidas en un elemento o en un grupo de elementos, definidos previamente como puntos de acceso. El lenguaje de consulta permite truncar por la derecha, izquierda y en medio, la utilización de operadores booleanos, de relación y proximidad, etc.

Desde los registros recuperados en cualquier vista se puede navegar a los registros relacionados con ellos, pertenezcan o no al mismo ente.

Por lo tanto, y a la vista de lo definido anteriormente, tendría que tenerse en cuenta los siguientes aspectos en la recuperación del ente CDU:

- 1) En la vista alfabética se presentan ordenados los registros de CDU, representados por aquella información que se considera más significativa: la notación. Es una lista de notaciones.

El problema fundamental que hay que resolver en lo que se refiere al acceso alfabético es el de la ordenación. En ARIADNA se han asignado unos "pesos" a cada uno de los caracteres que constituyen el juego de la Biblioteca Nacional (basado en el CECP, *New Spanish Set*). Estos pesos van a ser siempre los mismos, sea cual sea la información que haya que ordenar:

- signos como ':' '-' o '/' ordenan como un blanco;
- a continuación ordenan signos como '(' '+' '=';
- después los números;
- por último se ordenan las letras;
- signos como comillas "" o apóstrofo se convierten a nulo, es decir, desaparecen a efectos de ordenación y se ordena por el carácter que va detrás de ellos.

Como es fácil comprobar, la ordenación que resulta de la aplicación de estos pesos difiere de la que se recomienda para las notaciones de CDU: '+' '/' número ':' '=' '(' "" letra '-' '.0'.

- 2) En las vistas de recuperación por palabras se debe poder recuperar sin redundancias cada uno de los elementos que constituyen la notación. Además, se debe poder recuperar las definiciones asociadas a cada notación básica. De este modo, el usuario tendrá acceso a la notación a partir del término expresado en lenguaje natural, por un camino y con unos términos distintos de los que representan los encabezamientos de materia.

Por último, debe recordarse que se puede recuperar una notación de CDU mediante la navegación desde otro tipo de registro: autoridades de materia, elemento 080 del registro bibliográfico y los elementos 852h, 852i y 852k del registro de fondo

Evolución y futuro de la recuperación por la CDU

En general, el uso de sistemas de clasificación en catálogos en línea, como afirman Songqiao Liu y E. Svenonius (2), es todavía asunto de discusión y experimentación, pero estos experimentos y diseños de nuevos modelos son cada vez más numerosos ya que resultan evidentes las enormes posibilidades que ofrecen.

La principal diferencia entre los diferentes sistemas de clasificación y las estructuras sistemáticas basadas en vocabularios controlados es que cuando se asigna a una obra un número de la CDU no solo se le agrupa con obras similares sino que se le sitúa en un sistema jerarquizado con materias relacionadas. En dos esclarecedores trabajos de Roy Larson (3, 4) demuestra hasta qué punto van avan-

zando procedimientos que permiten identificar hasta un 80% de los nuevos registros correctamente clasificados.

Gran parte de estos proyectos están basados en la técnica de relacionar registros bibliográficos con términos adicionales, tales como ficheros de autoridades de materia o tesauros externos a la base de datos. El hecho de que los registros de autoridad de clasificación no estén preparados en la actualidad ni sean operativos en ninguna base de datos no debe hacernos perder de vista el hecho del análisis o procedimiento de análisis fuertemente matematizado propuesto por el autor. En todo caso debe quedar claro el hecho de que un número muy significativo de búsquedas, entre el 30 y el 45%, no consigue un resultado satisfactorio.

En cuanto al futuro de los distintos sistemas de información resulta particularmente interesante el trabajo presentado por N. J. Belkin (5) en el cual se citan cuatro diferentes proyectos de interfaz llevados a cabo actualmente en la Agencia Espacial Europea. A nuestros propósitos nos interesa recalcar el hecho de que los autores insisten en la necesidad del acceso a los distintos niveles basados en el análisis multinivel de los objetivos de los usuarios, sus distintas tareas y, en último término, dominios.

También tratando del diseño de las próximas generaciones de catálogos en línea Charles R. Hildreth (6) presenta cinco corolarios que son particularmente interesantes. El primero es suponer que el usuario va a poder navegar sobre bases de datos cada vez más potentes sin que exista una ayuda cada vez más eficaz de forma interactiva. Nunca se debe permitir que una búsqueda fracase (recordemos el 45% que cita Larson) y que la respuesta sea 0. La base de datos debe contener registros relacionados sistemáticamente con la búsqueda fracasada por autor. Muy importante es la recomendación de Hildreth de no aceptar jamás que la presentación de un registro al final de una búsqueda sea el fin de la sesión sino que (como ya está operativo en ARIADNA con las materias) debe ser el punto de arranque para nuevas búsquedas.

Por ello, es especialmente interesante el informe preparado por la Biblioteca del Congreso (7). Como resultado de este análisis, que tal vez el tiempo pueda hacer revisar, el autor recomienda el estudio de la aplicación de los sistemas expertos en el proceso de clasificación bibliográfica. Así, en la página 30 de su informe, se repasa la posibilidad de crear un *subject cataloguing consultant*, que incluya la clasificación.

Por último, y en lo que se refiere a perspectivas de futuro, en el anexo B de la norma ISO/DIS 10163 (8) se cita, en la página 42, entre los atributos hasta ocho tipos de distintos dedicados a la clasificación. Así, el *AttributeType—of—term, AttributeValue 13,—Classification, UDC* lo que asegura la perdurabilidad de la clasificación sistemática a la hora de proceder a realizar búsquedas entre sistemas abiertos.

Es evidente que el modelo ensayo/error puede proporcionar, si se cuenta con la suficiente flexibilidad, un acercamiento cada vez mayor al objetivo deseado. Para ello es necesario contar con un procedimiento de toma de decisiones que permitan modificar el análisis propiamente dicho como posteriormente la estructura de la información, y en su caso, la programación misma. Pero, para llevar a cabo la toma de decisiones que supongan la modificación de un sistema, es necesario contar con una información que no solo sea cualitativa, sino fundamentalmente cuantitativa.

Ya en un apartado anterior de este mismo trabajo nos hemos referido a la creación de los registros y a las asociaciones que se hayan de establecer entre ellos tanto en una fase inicial de carga masiva como en la labor diaria de actualización del fichero de formas utilizadas de CDU.

Mucho más importante es la valoración de los resultados de las consultas realizadas por los usuarios. El sistema de información debe ser capaz, como ocurre efectivamente en ARIADNA, de escribir una traza cada vez que se produzca un acceso a las distintas ocurrencias de la vista de registros de autoridad CDU tanto si esta se produce en los accesos alfabéticos como en los accesos textuales tan importantes, como es obvio, en el caso de los sistemas de clasificación sistemática.

Por otra parte, el módulo de OPAC, y asimismo por el procedimiento de grabar una traza cada vez que se produzca una consulta en la opción sistemas de clasificación, proporcionará la valoración ajustada de la utilidad de este procedimiento e, incluso, permitirá orientar la línea de futuras políticas generales y de indización de la biblioteca. Es muy importante señalar que las sucesivas navegaciones, es decir, los sucesivos pasos del modelo conceptual al modelo real y viceversa que se produzcan en

el OPAC deberán reflejar, asimismo, la oportuna traza, de tal manera que sea posible evaluar la pertinencia de la consulta y orientar futuros desarrollos que, fundamentalmente, se dirigirán, como parece probable, a través de las asociaciones TE y TG que permitirán la explotación de la estructura jerárquica de la CDU.

Referencias

- 1) Agenjo, Xavier. La automatización de Biblioteca Nacional recapitulación histórica. *Boletín de la Asociación Española de Archiveros, Bibliotecarios, Museólogos y Documentalistas* [en prensa].
- 2) DORS. DDC. On line Retrieval Systems. *Library Resources & Technical Services* 35(4):359-375, 1991.
- 3) Experiments in automatic library of Congress Classification. *Journal of American Society for Information Sciences* 43(2):130-148, 1992.
- 4) Evaluation of advanced retrieval techniques in an experimental online catalog. *Journal of the American Society for Information Sciences*. 43(1):34-53, 1992.
- 5) User interfaces for information systems. *Journal for Information Sciences*. 17(1991):327-344.
- 6) Beyond boolean: designing the next generation of line catalogs. *Library Trends*. Spring 1987.
- 7) Fenly, Charles. Expert systems. Washington: Cataloging Distribution Service, Library of Congress, 1988.- 37 p.- . (Advances in library information technology, ISSN 0899-1227; issue No. 1).
- 8) Documentation-Search and retrieve protocol specification. [Geneve?], International Standardizing Organization, 1990.- [50] p.

Recibido: 20 de diciembre de 1992.

Aprobado: 23 de febrero de 1993.

Xavier Agenjo
Biblioteca Nacional
Paseo de Recoletos, 20
28001 Madrid
España

PROCI

DIRECTORIO



DIRECTORIO EXPANDIDO DE PROFESIONALES EN CIENCIAS DE LA INFORMACION

El directorio PROCI es un producto elaborado por el Centro de Estudios y Desarrollo Profesional en Ciencias de la Información (PROINFO), te ofrece información detallada sobre especialistas en Ciencias de la Información que trabajan en diferentes partes del mundo.

PROCI brinda información sobre los aspectos relacionados con el grado de especialización, sus temas de interés, formas de contactarlos, sus publicaciones y otros datos personales.

Los datos se ofrecen en español e inglés, lo que permite su utilización por un amplio círculo de interesados.

PROCI está dirigido a los profesionales y técnicos que laboran en instituciones de información, bibliotecas y centros de referencia.

Si usted desea ser incluido en PROCI, puede dirigirse a PROINFO para solicitar un modelo de ficha descriptiva. La inclusión en PROCI es gratuita.

Para mayor información puede dirigirse a:

PROINFO 

I N G L E S / E S P A Ñ O L