

La organización de la información, los lenguajes documentales y la normalización.

Inocencia Soria González

Consejo Superior de Investigaciones Científicas .

Unidad de Coordinación de Bibliotecas

C/ Jorge Manrique, 27.

28006 Madrid

tel : (91) 585 49 83

fax : (91) 564 42 02

e-mail : soria@olivo.csic.es

1.La información de calidad

Desde mediados de siglo venimos asistiendo a un cada vez mayor crecimiento de la información propiciado en gran medida por el desarrollo de nuevas tecnologías aplicadas a la edición, la informática y la telemática y la aparición de nuevos soportes.

A esta abundancia de información se ha venido a sumar en los últimos años el fenómeno Internet.

Cualquier persona que disponga de un equipo no excesivamente complejo tiene la posibilidad no solo de acceder a información remota sino de introducir información (de calidad o no) de forma sencilla, aportando su granito de arena para contribuir a esta explosión informativa

Estamos, pues, rodeados de información potencial, tenemos una gran cantidad de datos, pero eso no significa que nuestras necesidades estén cubiertas. A menudo lo poco organizada que está, lo efímero de su validez, y el convencimiento o la sospecha de que no toda tiene la calidad que necesitamos, provoca que el proceso de búsqueda de información se convierta en una tarea difícil que consume más tiempo del que en principio habíamos imaginado.

Actualmente nuestro problema no es la cantidad sino la calidad de la información. Justamente por ello, las bibliotecas, el buen oficio y la experiencia de los bibliotecarios pueden jugar un importante papel en el suministro de información de calidad.

Quizás el hecho de proporcionar información de calidad está más relacionado con las tradicionales tareas bibliotecarias de lo que suponíamos, especialmente con la selección de la colección (real o virtual) y con el amplio conocimiento de los recursos y de las fuentes de información de reconocido prestigio.

Entre los factores que influyen en la efectividad de la recuperación de información, unos están relacionados con las características de los medios técnicos que utilicemos (capacidad, potencia, flexibilidad); otros, con la experiencia y conocimiento del funcionamiento de los mecanismos de que dispone ese medio y la familiarización con los programas (tipo de operadores, posibilidad de truncar términos, etc.); otros, con el acierto o no en el modo de expresar las necesidades de información ; y otros, no menos importantes, con la calidad en el tratamiento de la información del recurso (catálogo, base de datos, directorio, etc.) en el que efectuamos la búsqueda.

La urgencia de poner a disposición del usuario la información lo más rápidamente posible ha conducido a que algunos profesionales defiendan un tratamiento breve o superficial en el análisis y descripción de los documentos.

El entusiasmo que despertó en el mundo bibliotecario el hecho de la automatización de algunos procesos y las posibilidades que ofrece el hipertexto y los potentes mecanismos automatizados de búsqueda nos hizo concebir la esperanza de que podrían suprimirse algunas tareas, precisamente aquellas más engorrosas, que exigen determinados conocimientos especializados, rigor, experiencia y minuciosidad.

Pasado el primer espejismo, comprobado que tenemos miles de bytes que para todo propósito práctico son inexistentes y con el convencimiento de que ninguna cantidad de cables nos dará el acceso universal a las publicaciones volvemos a la creencia de que no hay una forma mágica, sin esfuerzo, de organizar la información y a afirmarnos en la idea de que donde quiera que ésta se produzca se necesita alguien que la someta al consiguiente tratamiento para hacerla asimilable y accesible. El proceso que

sirve de base a las distintas tareas de este tratamiento es conocido en el mundo bibliotecario con el nombre de control de autoridades.

2.-El control de autoridades

El principio básico y fundamental del control de autoridades es establecer una única forma autorizada para referirse a los datos bibliográficos (autor, título uniforme, serie, materia) que son potencialmente comunes para más de un documento. Las palabras usadas para expresar esa única forma autorizada o el orden en el que esas palabras aparecen es de consideración secundaria.

Una vez decidida esta forma única, se elabora un registro de autoridad estableciendo referencias desde otras formas de expresión bajo las cuales podrían aparecer esos mismos datos, para remitir a la forma autorizada y se citan las fuentes consultadas que justifican esa decisión. El registro así elaborado pasa a formar parte del fichero de autoridades.

Cuando el nombre, materia o título aparece en un documento con una forma diferente de la admitida, la forma autorizada prevalece sobre la forma que se encuentra en el documento, de modo que éste se incluye en el catálogo bajo el punto de acceso de la forma controlada.

Este sistema puede parecer caro y complicado, pero es la única manera de proporcionar con certeza la recuperación de todos los documentos de una colección que tienen un elemento en común: el mismo autor, materia...

Esto era una verdad indiscutible en la época de los catálogos manuales, constituidos como una secuencia alfabética. Las posibilidades de búsqueda eran muy limitadas. Si no se establecían los puntos de acceso apropiados y se ordenaban los registros bibliográficos, éstos se dispersaban dentro del catálogo y se perdían. Los catálogos automatizados no tienen esa limitación. Podemos recuperar un documento a partir solo de un fragmento de información que conocemos: una parte del nombre del autor, etc. Dada esta capacidad del catálogo automatizado ¿qué necesidad hay de controlar la forma de tantos datos?

Sin tomar en consideración que la automatización no ha llegado absolutamente a todas las bibliotecas, el argumento fundamental para el control de autoridades sigue teniendo sentido : a pesar de que los ordenadores pueden buscar datos a una velocidad enorme, sólo recuperarán los registros que contengan los datos concretos por los cuales han sido preguntados.

Paradójicamente fue la llegada del catálogo en línea, con el cual muchos creyeron superada la necesidad de control de autoridades, lo que dio a esta tarea un nuevo impulso. Los programas de gestión de catálogos facilitaban la elaboración de ficheros de autoridad, permitiendo establecer relaciones entre las distintas formas admitidas y no admitidas y asignar mecánicamente a cada registro bibliográfico individual la forma adecuada como punto de acceso. Además ofrecían la posibilidad de que, si se decidía cambiar la forma del encabezamiento, no era necesario cambiarla en cada registro individual, se podía cambiar en el fichero de autoridades y automáticamente resultaba corregida en cada registro.

Comprobado que cuanto más extensa es la información a manejar, más son los problemas causados por la falta de control de autoridades, hoy prácticamente nadie discute la importancia de esta actividad.

El establecimiento de un control de autoridades sólido y consistente exige la utilización de normas bibliográficas adecuadas.

3. La normalización

En cualquier actividad la normalización permite simplificar y racionalizar los métodos y las técnicas y unificar los productos. En el mundo bibliotecario la normalización ha sido la clave para permitir la transferencia de la información más allá de los límites nacionales, acelerar el proceso de su tratamiento y disminuir su coste.

Las organizaciones que están implicadas con el desarrollo de normas son muy eficientes cuando desarrollan temas tan esenciales y prácticos como medidas, tallas, tensiones, fuerzas, velocidad de transportes aéreos, etc.

Por el contrario, tanto la elaboración como la aprobación y aceptación de las normas bibliográficas ha tenido un desarrollo lento, en parte porque ya existían muchas normas prácticas nacionales establecidas tradicionalmente y también por el hecho básico de que en materia de documentación no existía la urgente necesidad, como en la industria, de efectuar cambios para vender los productos.

Con los comienzos de la informatización de la Library of Congress a mediados de los años sesenta y las posibilidades que ese hecho conllevaba, se fue desarrollando el convencimiento de que la elaboración de normas bibliográficas y su aceptación era vital para los intereses de la economía (una vez, un solo registro) y el intercambio internacional (un registro para el uso de todos). Desde entonces se han alcanzado algunos éxitos, siempre con la vista puesta en las posibilidades de la informatización y las

necesidades de buscar soluciones internacionales.

La normalización en el ámbito bibliotecario entraña dificultades mayores que en otros campos. Estamos tratando con normas y reglas usadas para describir y regular el producto de muy variables elementos: la actividad intelectual del ser humano.

A esta dificultad "per se" vienen a añadirse hoy dificultades nuevas. Si siempre hubo problemas con las obras anónimas, autores que adoptan diversos nombres, autores que utilizan seudónimos para buscar el anonimato, publicaciones periódicas escindidas o títulos genéricos ; ahora, con un incremento grande en el número de autores publicados, existen estos y otros muchos problemas. Abundan las publicaciones realizadas por autores colectivos, cada vez más es necesario reconocer títulos que han sido traducidos, transliterados, alargados, acortados, cambiados. Hoy además tenemos que enfrentarnos con nuevos vínculos bibliográficos en la relación entre autor y obra, con problemas para identificar el lugar de publicación , con la necesidad de describir publicaciones que aparecen en soportes muy diferentes del libro...

Son muchas las normas relacionadas con el mundo de la informática que están actualmente intentando contribuir a suprimir las muchas barreras en el intercambio de la información que ha originado el hecho de que diferentes empresas impusieran sus sistemas para cautivar a los consumidores, especialmente las que se refieren a compatibilidad de sistemas informáticos, lenguajes de programación y sistemas operativos que conllevan la necesidad de conocer diferentes protocolos y comandos, juego de caracteres, etc.

Pero aunque los sistemas informáticos sean compatibles, el intercambio no será posible si la información aparece de forma dispar, la descripción de los documentos no sigue los mismos métodos y normas o los campos en los que se registra la información no son los mismos.

Me limitaré exclusivamente a comentar las normas que afectan a la elaboración de registros bibliográficos.

Tradicionalmente en los catálogos manuales el registro bibliográfico se ha dividido en tres partes: encabezamiento principal, descripción formal y registro de encabezamientos secundarios que incluía la clasificación y la indización.

En la descripción formal es donde se ha conseguido mayor nivel de normalización y lógicamente en la clasificación y la indización donde menos.

Las normas que intervienen a la hora de elaborar registros bibliográficos son : las ISBD que afectan a la descripción formal ; las reglas de catalogación nacionales, basadas en las AACR, que afectan a la elección y asignación de los puntos de acceso ; y el formato MARC que afecta a la codificación electrónica de los datos del registro.

Con arreglo a las ISBD la descripción se estructura en ocho zonas, cada una de las cuales comprende unos elementos propios separados por signos de puntuación muy específicos. Cada área se obtiene de unas fuentes principales de información y tanto éstas como sus elementos son fijos en su número y orden de citación. La descripción puede hacerse a tres distintos niveles de detalle. Su exhaustividad dependerá de los propósitos del centro catalogador.

Aparte de incluir las ISBD, las reglas nacionales de catalogación se completan con normas para determinar los puntos de acceso principales y los manuales de instrucciones y listas para redactar nombres de persona o entidades y títulos uniformes.

Tradicionalmente el primer paso en la catalogación consiste o consistía en determinar el encabezamiento de la ficha principal o asiento de entrada atendiendo a estas reglas.

El formato MARC, diseñado para facilitar el intercambio de información bibliográfica legible por ordenador, codifica los datos del registro. Cada uno de los elementos de la ficha catalográfica tiene asignado un identificador numérico de tres cifras. Dentro de cada campo cuenta con indicadores que especifican su contenido, códigos de subcampo que identifican un elemento concreto del campo y delimitadores que señalan la localización del campo dentro del registro.

Provenientes del mundo editorial, la identificación de las publicaciones mediante un código numérico, tiene tanto un carácter bibliográfico como comercial. El ISBN y el ISSN permiten identificar una publicación en cualquier país y en cualquier lengua y por tanto facilitan operaciones de canje, préstamos, intercambio de datos, etc. a la vez que mejoran los circuitos de venta identificando rápidamente los pedidos entre editores, librerías y distribuidores.

Especialmente útiles han resultado estos códigos en la captación de registros en tareas de conversión

retrospectiva, siempre, claro está, que se trate de ejemplares editados con posterioridad a su instauración (la obligatoriedad de su constancia en la descripción en España no se hace efectiva hasta las reglas de catalogación de 1985)

4. Cuestionamiento de las normas bibliográficas

La funcionalidad de muchos aspectos de estas normas está siendo puesta en cuestión por muchos profesionales. La crítica general es que se están siguiendo normas y criterios propios de los catálogos manuales para realizar los procesos automatizados, que esas normas deben ser adaptadas a la nueva situación y que tienen que ser simplificadas.

Las críticas a las ISBD están dirigidas contra su inflexibilidad y conservadurismo que parecen querer perpetuar el formato de registro bibliográfico en ficha impresa, además de que la minuciosidad en los detalles sobre el orden de los elementos y la puntuación son innecesarios en un catálogo automatizado. Pese a estas objeciones, prácticamente todos los sistemas integrados de gestión bibliotecaria incorporan un formato de salida en ISBD.

Se plantean también críticas a la relación de la norma ISBD con las reglas nacionales de catalogación porque duplica la información en el mismo asiento al repetirse datos en la descripción y en los encabezamientos o puntos de acceso al documento. Como solución algunos expertos sugieren la supresión de los elementos informativos que son susceptibles de ser debidamente tratados y normalizados por el control de autoridades y que ya se hacen constar en el bloque de puntos de acceso al documento; esto es, todas las menciones de responsabilidad sin excepción, tanto personales como colectivas, los títulos uniformes y las series.

Como contrapartida se puede objetar que si se simplifica la descripción se pierde información. Si se suprimen las menciones de responsabilidad se omiten también las palabras que expresan el vínculo intelectual entre autor y obra aunque tal vez la información que se pierde no es sumamente relevante y se puede suplir mediante los códigos de función que acompañan a los puntos de acceso.

Las reglas de catalogación son las normas más directamente afectadas por la automatización. Su principio básico es la distinción entre encabezamientos principales y secundarios. Su función primordial, proporcionar una ordenación alfabética a las fichas, no tiene sentido en un catálogo en línea toda vez que en éste los asientos no reciben a priori ordenación alguna y la información que contienen es recuperada de modo eficaz, independientemente de la posición que ocupe en dicho asiento. Si en un entorno manual la información aparece en las distintas partes del catálogo ordenada bajo el apellido del autor, el título y éstos eran los elementos que constituían los puntos de acceso, en el catálogo electrónico los puntos de acceso se multiplican, ya que no es la disposición de la información sino las instrucciones de búsqueda las que permiten la recuperación.

El formato MARC, a diferencia de las otras normas, se diseñó para ser usado en catálogos automatizados. Pese a esto se le critica que, aunque su estructura posibilita la creación de registros legibles por ordenador con un sistema de campos y subcampos, desde el punto de vista conceptual sigue estando basado en esquemas, especificaciones y principios metodológicos propios del entorno del trabajo manual.

El formato MARC puede verse como una réplica mecanizada de la ficha catalográfica impresa de la que hereda: su división en tres partes (encabezamiento principal, descripción y encabezamientos secundarios); sus elementos más característicos (cada área ISBD tiene su correspondencia exacta en un campo MARC); y sus normas de base (ISBD y reglas de catalogación).

En la práctica el formato MARC contribuye a la duplicidad información. Algunos plantean como solución una simplificación del formato que suprima los campos que repiten información.

5. Los lenguajes documentales

Junto a la descripción formal, que permite la identificación de los documentos a partir del nombre del autor o autores o del título, el otro aspecto más destacado en la recuperación de la información es el relacionado con el acceso a los documentos atendiendo a su contenido. Diversos estudios demuestran que actualmente la búsqueda por temas es la más utilizada, especialmente en el campo de las ciencias, donde sobrepasa el 60% de las búsquedas realizadas.

En el marco de la recuperación temática hay que añadir a las dificultades propias de la normalización bibliográfica, las específicas de la naturaleza del lenguaje.

El lenguaje del discurso común está constituido por un número ilimitado de términos. Su propia riqueza comunicativa que permite gran variedad de matices se convierte en un impedimento a la hora del tratamiento de la información.

El lenguaje documental es el sistema convencional creado para poder expresar el contenido de los documentos sin los impedimentos del lenguaje natural. Simplifica el lenguaje común utilizando sólo una

pequeña parte del léxico, algunas formas y poco o casi nada de gramática.

Los lenguajes documentales pueden consistir simplemente en una lista de palabras admitidas aunque lo más frecuente es que consten de un sistema estructurado que relacione sus distintos términos.

Estos sistemas reducen las ambigüedades semánticas (diferenciando los distintos significados de los homónimos, controlando los sinónimos por medio de reenvíos) y ayudan además a encontrar, a través del entramado de su estructura de relaciones, temas cercanos (más amplios o más específicos) al objeto de la búsqueda.

Es precisamente en la recuperación por temas donde la automatización ha creado más inquietudes, en parte por las posibilidades que ofrecen los ficheros de palabras generados automáticamente a partir de cualquier campo del registro y en parte por las nuevas perspectivas que se abren con los experimentos de indización automática.

Los sistemas sin lenguaje controlado que se sirven de las palabras tal y como aparecen en el título, resúmenes o texto íntegro pueden ser útiles en los casos en que se supone que el lenguaje natural empleado es suficientemente preciso lo que ocurre con los lenguajes científicos y técnicos pero esta precisión es muchas veces más aparente que real y prácticamente inexistente en el caso de las ciencias sociales.

Los tipos de lenguajes documentales suelen ser divididos en los manuales profesionales de la siguiente manera :

Por el grado de control se dividen en lenguajes libres y controlados. En realidad lo que llamamos lenguaje libre no puede ser considerado como un lenguaje documental, es más bien una disposición u ordenación de las palabras más significativas de los títulos : índices KWIC y KWOC.

Por el momento de coordinación de los elementos del lenguaje en precoordinados (clasificaciones y encabezamientos de materia) y postcoordinados (descriptores) ; dependiendo de si se coordinan en el momento de la indización, los primeros, o en el momento de la búsqueda, los segundos. Los precoordinados son más rígidos, los postcoordinados más flexibles pero pueden dar lugar a coordinaciones erróneas .

Por el modo de su estructura podemos dividirlos en lenguajes de estructura jerárquica y combinatoria : los de estructura jerárquica organizan los conocimientos en un sistema de clases y subclases siguiendo una determinada lógica. Describen el tema principal del documento de forma sintética , y, a veces, uno o dos temas secundarios. Para ello se sirven de una codificación artificial, generalmente numérica.. Los lenguajes de estructura combinatoria permiten definir el contenido de un documento mediante un conjunto de palabras que expresan el tema o los temas de que trata.

6. Los Lenguajes de estructura jerárquica : las clasificaciones

Los lenguajes de estructura jerárquica tradicionalmente usados en las bibliotecas son las grandes clasificaciones universales.

Tienen un fuerte componente utópico: pretenden constituirse en instrumentos de organización global del saber humano y por lo tanto son empresas cuya magnitud exige la cooperación de muchos expertos.

Se critica su rigidez y dificultad de actualización y se alaba el hecho de que permiten una visión general en la que se puede profundizar con pocas posibilidades de error.

Los sistemas clasificatorios más difundidos son : la Clasificación decimal de Dewey (DDC), la Clasificación decimal Universal (CDU) y la Clasificación de la Library of Congress (LCC).

La LCC fue creada a fines del siglo pasado para encontrar una solución al fondo concreto de la LC cuando se trasladó a sus nuevos locales, un siglo más tarde la fuerza de atracción estadounidense ha conseguido que la LCC sea utilizada por miles de bibliotecas del mundo anglófono.

La DDC es usada en las bibliografías nacionales de muchos países anglosajones y ha sido traducida a varias lenguas. Su difusión se debe en parte a que es asignada a los registros MARC producidos por la Library of Congress. Es actualizada regularmente, no es que efectúe cambios radicales en cada edición pero, de las tres, es la que hace más esfuerzos para no quedarse obsoleta.

Me centraré un poco más en la CDU, que es el sistema que me resulta más familiar por ser el más difundido en España, pero casi todo lo que diga de él podría ser aplicado a la DDC.

La CDU es el sistema más utilizado en los países no anglófonos y es usada por un sinnúmero de bibliografías nacionales de todo el mundo. Sus más rendidos admiradores afirman que tiene más flexibilidad y más capacidad de expresión que la DDC y que no ha alcanzado el reconocimiento que

merece a causa de que la falta de fondos y el sistema burocrático de la FID, entidad encargada de su gestión hasta hace cuatro años, han dificultado sus revisiones ágiles .

En 1992 se constituyó el consorcio CDU, que asumió las responsabilidades que antes tuviera la FID sobre su edición, actualización, versiones, etc. Este consorcio, cuyos socios fundadores son Bélgica, España, Holanda, Japón, Reino Unido y la propia FID, se comprometió a organizar y mantener la CDU y sus esfuerzos ya han dado algunos importantes frutos : se ha creado un fichero informático con más de 60.000 entradas que está sirviendo de base para facilitar su manejo y actualización.

Aparte de su utilidad como elemento ordenador de los fondos, especialmente en salas de libre acceso para orientar al lector en su búsqueda por los estantes, el uso durante muchos años de los sistemas clasificatorios en las bibliotecas (la CDU ha sido obligatoria en España por la Ley de Bibliotecas durante 50 años) ha dado lugar a extensos catálogos sistemáticos que han pasado, por supuesto, a formar parte de sus catálogos automatizados.

El empleo de lenguajes clasificatorios en sistemas informatizados es de gran utilidad, según prueban numerosos estudios, sobre todo en combinación con otros sistemas de indización. El problema es que muchas de las aplicaciones informáticas no permiten aprovechar todo su valor.

Un problema frecuente es la necesidad de incluir en la recuperación términos relacionados jerárquicamente. Por ejemplo, alguien interesado en el transporte por tren podría querer también información relacionada con la señalización de vías, estaciones, etc. y resulta muy tedioso tener que pensar en todas las palabras por las que preguntar y unir en una búsqueda.

En algunos ficheros no existe, además, ningún otro campo de recuperación temática que pueda ser utilizado : bastantes ficheros de publicaciones periódicas carecen de encabezamientos de materia o de descriptores.

Algunos de los inconvenientes más señalados en la CDU -por ejemplo, el hecho de que la disposición jerárquica de sus números obliga, en principio, a realizar la búsqueda expresando toda la secuencia completa- podrían ser minimizados si los sistemas de gestión desarrollaran las herramientas informáticas adecuadas : si se pudieran usar en la búsqueda los operadores de truncamiento y los signos propios de la CDU (paréntesis, barra, dos puntos, signo más, etc) se aprovecharían los números auxiliares independientemente del número principal.

Durante el primer periodo del desarrollo de los servicios de información en red, muchos especialistas cuestionaron tanto el valor de los sistemas de descripción por materias, como de los sistemas de clasificación, argumentando que éstos podían ser cumplimentados por los programas de indización de texto completo.

El incremento del uso de Internet y WWW ha ido, sin embargo, cambiando esta percepción. Existen dos métodos principales de recuperación: uno es el de los robots que buscan por las palabras claves que contienen los WWW, muy útiles aunque suelen recuperar también extensas cantidades de información irrelevante ; el otro utiliza sistemas clasificatorios para organizar una estructura por índices que da acceso a sus recursos seleccionados. El uso de esquemas clasificatorios permite hacer más amplias o más específicas las búsquedas a la vez que permite un acceso multilingüe .

Algunos servicios, como Yahoo, tienen su propia clasificación, pero muchos otros, que dan acceso a recursos considerados como de buena calidad para un área concreta, están utilizando la DDC o la CDU. Una lista de los servicios que usan sistemas de clasificación puede encontrarse en <http://www.iastate.edu>.

Entre las ventajas señaladas por los servicios que los utilizan cabe señalar las siguientes: cubren todas las áreas temáticas, han sido revisados y actualizados por las entidades internacionales responsables, los usuarios que utilizan regularmente las bibliotecas están familiarizados con al menos una parte de su esquema, tienen buenas traducciones a muchas lenguas, están disponibles en formatos legibles por ordenador y pueden realizarse conversiones automáticas de unos sistemas a otros.

Las objeciones más extendidas se apoyan en que estos esquemas imponen un falso orden sobre el conocimiento; son subjetivos, inconsistentes, ilógicos y tienen una especial dificultad en incorporar con radidez nuevas áreas del conocimiento.

Uno de los ejemplos más significativos de la utilidad que puede tener la CDU en el entorno Internet es NISS, una guía temática que incluye recursos fundamentalmente británicos (<http://www.ac.uk>) NISS usa los números de la CDU como un marco para presentar la información por materias. El usuario no necesita conocer la CDU que es utilizada como una herramienta invisible para organizar la información.

La consulta puede realizarse fácilmente a través directamente de la jerarquía temática, por medio de un índice alfabético que conduce a la estructura de la CDU o por las palabras claves pertenecientes a un registro que describe cada recurso. Una vez que el usuario ha encontrado un recurso de su interés tiene la opción de conectarse directamente o de consultar el registro en el que se describe éste y averiguar

previamente si le interesa o no, antes de establecer la conexión .

7. Los lenguajes de estructura combinatoria : los encabezamientos de materia

Los sistemas de estructura combinatoria más utilizados para expresar el contenido de los documentos son los encabezamientos de materia y los descriptores.

Los encabezamientos de materia tienen una estructura precoordinada, es decir, se combinan en el momento de la indización mientras los descriptores se coordinan en el momento de la búsqueda.

Tradicionalmente los descriptores han sido utilizados en el tratamiento analítico, especialmente en el análisis de artículos de publicaciones periódicas, expresaban distintos aspectos del documento, y sus vocabularios característicos, los tesauros, eran la herramienta propia de los documentalistas, mientras los encabezamientos de materia expresaban de forma global el contenido de monografías y sus vocabularios, las listas de encabezamientos, eran la herramienta propia de los bibliotecarios.

Precisamente la precoordinación, que ha sido la ventaja fundamental para la recuperación temática en los catálogos manuales, pesa ahora como una losa en los encabezamientos y es el aspecto de que más contribuye a su rigidez y artificiosidad .

Las fronteras entre unos y otros se están difuminando bastante en la medida en que las listas de encabezamientos aparecen con una estructura de relaciones y una terminología propia de los tesauros y, además, los sistemas automatizados permiten la recuperación de información por palabras pertenecientes a cualquier campo del registro sin importar el orden en el que se hayan introducido en el sistema y por lo tanto evitan (en la búsqueda que no en la introducción de términos) los problemas del orden de encabezamientos y subencabezamientos.

A pesar de que los encabezamientos de materia son un sistema antiguo, desarrollado para los catálogos de fichas, tienen todavía mucho uso hoy en día, pese a las facilidades de búsqueda postcoordinada usando operadores booleanos que ofrece el catálogo automatizado.

Los encabezamientos de materia sobreviven por varias razones: la primera es que muchas bibliotecas todavía tienen catálogos en fichas o en microforma en los que la búsqueda depende de la precoordinación de los términos, la segunda es que cientos de registros relativamente baratos, impresos o legibles por máquina que contienen encabezamientos de materia de la Library of Congress (LCSH) están a disposición de muchas bibliotecas. Es un método rápido y barato, aunque su calidad no sea la que muchas bibliotecas desearían .

Además, aunque existiera idealmente el deseo de cambiar y muchos bibliotecarios fueran partidarios de un sistema nuevo de catalogación por materias, en la práctica el cambio depende de grandes agencias, como la LC, que manejan inabarcables BBDD y están limitadas por la escasez de presupuestos y los insuficientes recursos humanos. A la vista de esta situación es lógico concluir que, deseable o no, el uso de los encabezamientos de materia continuará en un futuro previsible.

Aunque los LCSH no son una norma sino el producto de la práctica y del día a día del trabajo de una gran biblioteca, en realidad han sido y son considerados como una norma a través de la cual otras bibliotecas determinan la calidad y nivel de su catalogación por materias.

Algunas bibliotecas han hecho de los LCSH una aplicación ilimitada y acrítica sin tener en cuenta los objetivos, dimensiones, tipo, fondos o materia a la que está dedicada, contrariamente a la primitiva filosofía de su fundador. Cutter concibió los encabezamientos como una serie de reglas y principios comunes que serían aplicados y administrados por las bibliotecas a medida de sus necesidades, de tal modo que cada cual desarrollaría su propia lista y aplicaría sus propias políticas e interpretaciones.

Valorada por su extensión y criticada tanto por sus lagunas como por la calidad de su terminología, la lista de encabezamientos de la LC intenta acercarse a un lenguaje más natural y menos artificioso, pero su actualización no es tarea fácil. Nunca ha habido una revisión completa del sistema, cada nueva edición es en realidad la edición previa incorporando los nuevos cambios o adiciones realizados desde que fue publicada la anterior.

No hay reglas precisas para la precoordinación u orden de citación de los términos. Perduran muchas complicadas estructuras diseñadas para ayudar a la ordenación alfabética en los catálogos de fichas pero que crean dificultades en los usuarios de los catálogos automatizados.

Algunas investigaciones realizadas en torno a las proposiciones de convertir los LCSH en un tesoro, aplicando la norma ISO-2788 y transformando un significativo número de encabezamientos precoordinados en descriptores, determinaron que la estructura de relaciones es tan deficiente que sería necesario crear una nueva y que el esfuerzo que necesitaría emplearse en este proyecto es de tal envergadura que debe ser desestimado.

Las bibliotecas que no son de habla inglesa deben hacer esfuerzos adicionales para aplicar los encabezamientos eficazmente. A causa de la naturaleza etnocentrista de la lista hay temas como la

política, literatura, pueblos nativos y cultura general que no están adecuadamente cubiertos para las colecciones no estadounidenses.

Pese a estos y otros defectos, en lo que atañe a nuestra experiencia en el catálogo colectivo de las bibliotecas del CSIC, la lista de la LC ha sido y sigue siendo el principal apoyo para la normalización. El fichero de autoridades de materia de la red de bibliotecas del CSIC, fruto de diez años de trabajo, contiene más de 40.000 entradas entre términos admitidos y no admitidos y está construido con el sistema de relaciones propio de los tesauros, al estilo de los LCSH.

Hace dos años publicamos este fichero impreso en papel añadiendo además un glosario inglés-español que contiene unos 20.000 términos correspondientes a los LCSH que consideramos equivalentes a otros tantos encabezamientos nuestros.

En nuestro trabajo hemos intentado huir de las traducciones literales, hemos incluido lógicamente encabezamientos relacionados con nuestra cultura, derecho, historia o instituciones que no están representados en los LCSH y hemos usado términos distintos en los casos en los que diferíamos de la interpretación de algunos conceptos.

Actualmente se están desarrollando algunas experiencias de uso de encabezamientos de materia en el marco de Internet : INFOMINE, una biblioteca virtual creada por la Universidad de California, que contiene más de 5.000 registros que describen y dan acceso a una cuidada selección de recursos académicos relacionados con la biología, agricultura y medicina, ha desarrollado un método para usar encabezamientos LCSH de una manera simplificada (<http://lib-www.ucr.edu>).

8. La indización automática

Si la indización es ya en sí misma una tarea compleja, su dificultad se multiplica cuando se intenta hacer de forma automática (el sistema se enfrenta a muchos problemas a la hora de analizar el lenguaje natural: anáforas, elipsis, siglas, números...)

En su consecución intervienen entre otras disciplinas, la lingüística, la informática, la estadística y la documentación.

Había y hay profesionales que sostienen que una máquina es incapaz de indizar porque no puede llegar a captar los matices conceptuales, otros opinan que no resuelve nada porque los términos deben ser luego revisados por el indizador humano y otros que sólo puede ser aplicada a áreas del conocimiento muy concretas. Los defensores alegan mayor economía, mayor objetividad porque se aplicarían siempre los mismos parámetros y la disminución de los errores.

Las primeras investigaciones se llevaron a cabo en los años sesenta fundamentándose en principios estadísticos y probabilísticos, para dejar paso, diez años después, a métodos más complejos basados en el valor de discriminación y la relevancia de los términos. Actualmente se trabaja más con criterios lingüísticos a la hora de efectuar el análisis de los textos, si bien existen los sistemas híbridos. En casi todas las experiencias, una de las etapas consiste en la validación con un tesoro.

Hay poca presencia de bibliotecarios y documentalistas en estas investigaciones mientras abundan informáticos y lingüistas

9. Conclusiones

- La abundancia de información, que puede llegar a agobiar al usuario, exige a los bibliotecarios afinar mucho más, tanto en el conocimiento de las fuentes en las que se encuentra y en sus métodos de recuperación como en el escrupuloso tratamiento a la hora de procesarla.
- Debe darse un incremento en el número de puntos de acceso. La norma de limitar los puntos estaba concebida para no aumentar el volumen de los catálogos (uno de los objetivos de los catálogos manuales era incluir la mayor cantidad de información en poco espacio). Ahora no hay razón de espacio y las posibilidades cada vez más amplias de la tecnología nos obligan a plantearnos no solo si el catálogo debe permitir el acceso a los documentos completos sino también la profundización en el tratamiento documental que sea capaz de recuperar capítulos concretos de una obra, añadir sumarios, resúmenes, etc. Este planteamiento trae consigo un sinfín de preguntas: ¿en qué documentos sería útil?, ¿quién la haría : documentalistas, bibliotecarios?, ¿cuánto cuesta?
- El esquema de recuperación tradicional ha sido transportado a los sistemas automatizados. Es necesario que los programas de gestión de catálogos presten más atención al perfeccionamiento de los mecanismos de búsqueda. Algunos procesos del tratamiento documental contruidos tan laboriosamente resultan casi baldíos. Especialmente infrutilizados, como instrumento sugerente, están los sistemas de relaciones entre los distintos términos y las notas que aclaran su significado y delimitan su uso y alcance. Resulta a estas alturas inaudito que algunos sistemas respondan cuando se les pregunta por un término no admitido que no hay información, en lugar de establecer automáticamente la relación con el admitido o mostrar las referencias que expliciten otra alternativa más genérica o

relacionada. El ajuste más estudiado entre las herramientas informáticas y las bibliotecarias nos permitiría sacar mucho más rendimiento de los métodos de tratamiento de la información con los que contamos actualmente y que, pese a todos sus defectos y obsolescencia, son las que realmente conocemos, por lo menos hasta que seamos capaces de elaborar y comprobar los resultados de otros nuevos.

Bibliografía

COYLE, Karen.. **Acces : not just wires**. <<http://www.dla.ucop.edu/~kec/njw.html>> [consulta : 7 de octubre 1997]

GIL LEIVA, I. ; RODRIGUEZ MUÑOZ, J.V. "Tendencias en los sistemas de indización automática. Estudio evolutivo." **Revista Española de Documentación Científica**, 1996, vol 19, N°3

GONZALEZ, Fernando Edmundo. "La recuperación de la información automatizada :sus posibilidades para reorientar algunas funciones bibliotecarias". **Investigación Bibliotecológica**. Vol 4, N°8

GONZALEZ SERENO, Elvira ; SORIA GONZALEZ, Inocencia "Evaluación del uso de los encabezamientos de materia en el catálogo colectivo CIRBIC frente al uso de palabras de cualquier campo para la recuperación de la información" en. **V jornadas españolas de documentación automatizada**. Cáceres, 1996

LANCASTER, Frederick Wilfrid **El control del vocabulario en la recuperación de la información**. Valencia : Universitat de Valencia, 1995

MENENDEZ PELAYO, Jesús J ; BONACHERA CANO, Francisco J. "Propuesta para la optimización de asientos catalográficos en catálogos en línea de la primera generación". **Revista Española de Documentación Científica**, 1997, vol 20, N°1

Standards for the international exchange of bibliographic information. London : Library association, 1991

MITCHELL, Steve. **Library of Congress Subject Headings as subject terminology in a virtual library : The INFOMINE example**. <<http://lib-www.ucr.edu/pubs/postlcs.html>> [consulta : 7 de octubre de 1997]

NISS resource descriptions database guide. <<http://www.niss.ac.uk/resource-description>> [consulta : 7 de octubre de 1997]

The role of classification schemes in Internet resource description and discovery . <<http://www.ub2.lu.se/desire/radar/reports/D3.2.3/>> [consulta : 7 de octubre de 1997]

The role of classification schemes in Internet resource description and discovery . <<http://www.ub2.lu.se/desire/radar/reports/D3.2.3/>> [consulta : 7 de octubre de 1997]

VAQUERO, José Raúl. **Herramientas de búsqueda en Internet**. <<http://exlibris.usal.es/~rvaquero/teoria/lteo/htm>> [consulta : 7 de octubre de 1997]

