

RECENSIONES



MOYA ANEGÓN, F. de. *Los sistemas integrados de gestión bibliotecaria*. Madrid: ANABAD, 1995, ISBN 84-88716-15-X.

Aunque en los últimos años se está llevando a cabo un considerable esfuerzo por parte de varios editores españoles por aumentar y completar la bibliografía profesional en nuestra área de conocimiento –ya sea a través de traducciones o a través de la edición de originales–, ésta todavía sigue siendo insuficiente, detectándose importantes lagunas en algunas materias. En especial, resulta sorprendente la escasa presencia de manuales o monografías dedicadas a la recuperación de información por medios automatizados.

La Recuperación de Información es sin duda el área más importante en el seno de la Documentación, de la que fue su verdadero germen en los años 50. Esta no sería lo que es en la actualidad si no fuera por los trabajos de pioneros como Luhn, Taube o Mooers en los problemas de recuperación de información mediante procesos automáticos. A este respecto, hay dos trabajos recientes que demuestran de manera empírica tales afirmaciones. En el primero de ellos, llevado a cabo por los finlandeses Jarvelin y Vakkari (1), se estudian los temas tratados en las principales revistas de Bibliotecomanía y Documentación entre 1965 y 1985. Sus resultados demuestran que el área más frecuentemente tratada –y a considerable distancia de las demás– es la Recuperación de Información, que además muestra un considerable aumento en su porcentaje desde 1965 a 1985. Por lo que se refiere al segundo de ellos (2), se dedica a analizar la revista JASIS entre 1986 y 1990, con el objetivo de determinar cuáles son los frentes de investigación y cuáles son las personas que los encabezan. El estudio agrupa las áreas de investigación mediante el uso de clusters, de los que el correspondiente a Recuperación de Información es el de mayor tamaño, a gran distancia del resto. Además, y como consecuencia lógica de lo anterior, sus principales representantes (Salton, Van Rijsbergen, Croft, Bookstein, Sparck-Jones...) aparecen como los autores más citados en los artículos de JASIS.

Pues bien, precisamente es en Recuperación de Información donde se detectan las lagunas más importantes en nuestra bibliografía profesional, por lo que la mera aparición de la obra del Prof. Moya, que aborda en profundidad algunos de sus problemas básicos, debe ser valorada muy positivamente.

El origen de esta monografía –como el propio autor aclara– se encuentra en su tesis doctoral, lo que determina la estructura de la obra. Así, dejando aparte la introducción (cap. I), el primero de sus capítulos está dedicado a los problemas de carácter metodológico, los cuatro siguientes son el resultado del análisis de la bibliografía consultada y, finalmente, el séptimo y último consiste en relacionar las últimas investigaciones en la materia con diversas pruebas con datos reales llevadas a cabo por el autor. Esta estructura se ajusta perfectamente con el doble objetivo que éste persigue en la elaboración de la obra: a) analizar los modelos de estructuras de datos utilizados por los sistemas integrados de gestión bibliotecaria (SIGB), determinando la relación entre dichas estructuras y sus prestaciones en cuanto a recuperación de información (cap. III a VI); y b) examinar diversas técnicas avanzadas de recuperación de información, que todavía no han sido implementadas en los SIGB, para proponer una estructura de datos alternativa que permita mejorar las prestaciones que ofrecen los SIGB actuales (cap. VII). A este respecto, hay que señalar que el título original de la tesis es “Estructura de datos y recuperación de información”.

El capítulo II, dedicado a las cuestiones metodológicas, explica detalladamente las fases para la realización del trabajo. Comienza con una descripción de la biblioteca como estructura organizacional, utilizando para ello uno de los modelos clásicos de análisis de las organizaciones: el denominado "entidad-relación". A continuación, lleva a cabo su análisis funcional, usando tanto modelos de datos como de procesos, que le permitan establecer un marco conceptual sólido, con independencia de cualquier software o hardware concretos. La tercera y última fase tiene el objetivo de proponer soluciones específicas para la gestión de la biblioteca, lo que se lleva a cabo a través de cinco subfases: organización de datos en tablas y/o registros, organización de procesos, soluciones informáticas concretas, interface hombre/máquina y programación.

En el siguiente capítulo se aborda el análisis de la estructura funcional básica de la biblioteca, para lo que el autor ha tomado como punto de partida las cinco funciones básicas que se pueden distinguir en la estructura interna de cualquier biblioteca y que, por otro lado, coinciden con los módulos habituales de los SIGB: adquisiciones, catalogación, circulación, control de publicaciones periódicas y referencia. Para cada una de estas funciones básicas se describen las distintas operaciones que deben ser mecanizadas, de manera que se pueda definir la estructura de las correspondientes entidades y relaciones.

Los tres capítulos siguientes (IV a VI) se ocupan del análisis de los tres modelos de gestión de información utilizados por la informática: el modelo relacional, los sistemas de recuperación de información (IRS) y los de gestión de ficheros (FMS). Cada uno de ellos hizo su aparición en el mercado con una vocación claramente universalista, por lo que, al menos por un corto período de tiempo, fue el más aplicado en los sistemas bibliotecarios automatizados. El análisis pormenorizado de cada uno de ellos por parte del autor va poniendo de manifiesto sus respectivas ventajas e inconvenientes. Así, por ejemplo, el modelo relacional resulta muy adecuado para el tratamiento de información muy estructurada, no textual, con facilidad para la tabulación, en tanto que los IRS, por el contrario, tienen su punto fuerte en la recuperación de información textual. En cuanto a los FMS, aunque se les reconoce el acierto de ser los primeros que intentaron gestionar de manera integrada las bibliotecas, su estructura de datos no permite la utilización de técnicas de recuperación de cierta sofisticación, como las basadas en la similaridad, en la ponderación de las entradas o en procesos de feedback. También se les critica su excesiva dependencia de un hardware y software concretos y su escasa capacidad de adaptación, por lo que las aplicaciones basadas en este modelo han ido desapareciendo poco a poco.

La lectura de estos tres capítulos, además del dedicado a la estructura funcional de la biblioteca, nos proporciona una visión muy clara de cuáles son los requisitos que debe cumplir un SIGB para satisfacer de manera adecuada las necesidades de una biblioteca en la actualidad, tanto en lo que se refiere al trabajo que tienen que desarrollar en su seno los bibliotecarios como a los usuarios que pretenden acceder a sus fondos o utilizar sus servicios documentales. Nos permite, además, descubrir por qué los SIGB que hemos ido utilizando, ya sea como bibliotecarios o como simples usuarios, no nos ofrecían algunas posibilidades y prestaciones que nos parecían totalmente necesarias —o al menos deseables— cuando hacíamos uso de ellos. Sin duda, la razón de que estos capítulos resulten tan interesantes para cualquier persona atraída por la informática aplicada a la gestión bibliotecaria, estriba en la dilatada experiencia del autor en instalaciones de SIGBs en distintos contextos bibliotecarios, así como en su

colaboración en el desarrollo y diseño de alguno de ellos. Dicha experiencia, unida a su labor docente e investigadora en informática documental, le proporcionan una visión global y abarcadora, tanto a nivel teórico como práctico, que se refleja en la claridad con que analiza los problemas y propone posibles soluciones.

En el séptimo y último capítulo, titulado "Sistemas híbridos", el autor afronta la tarea de definir una estructura de datos apropiada que permita implantar en los SIGBs las técnicas más avanzadas de recuperación de información. Dicha estructura debería estar soportada por los denominados "sistemas híbridos", caracterizados por su naturaleza mixta, ya que intentan gestionar de manera integrada y eficaz tanto información textual como estructurada, superando las limitaciones respectivas de los DBMS e IRS.

Como labor previa a la definición de la estructura de datos, se lleva a cabo una completa revisión de cuáles son los principales problemas que presentan los métodos tradicionales de recuperación de información –de tipo "exact match" y basados en el álgebra de Boole– utilizados en los SIGBs. A continuación, se comienzan a describir qué elementos debe incluir un sistema que supere tales problemas, es decir, en qué consisten exactamente las técnicas avanzadas de recuperación de información. Estas tienen como punto de partida que el objetivo no es una equiparación exacta (exact match) entre los términos utilizados en la búsqueda y los utilizados en la indización, sino que se busca el mayor grado de similitud (partial match) entre tales términos, lo que permitirá con posterioridad establecer una ordenación de los resultados obtenidos en la búsqueda de acuerdo con su relevancia.

Las investigaciones acerca de cómo calcular el grado de dicha similitud –denominado "coeficiente de similaridad"– han seguido tres vías, tres modelos teóricos: probabilístico, de lógica difusa y de espacio vectorial. Este último se basa en que tanto las demandas como los documentos pueden representarse mediante vectores, cuyos componentes son los términos –con su correspondiente peso– utilizados en la búsqueda y en la indización, respectivamente. De esta forma, resulta muy fácil el cálculo de la similaridad mediante la comparación de los vectores correspondientes. Esta simplicidad de uso ha provocado que sea el modelo más extendido y desarrollado, por lo que ha sido el escogido por el autor.

Dentro de este modelo, una de las cuestiones más difíciles de resolver es la determinación automática del peso adjudicado a cada término de indización, también denominado "valor de discriminación". Para llevar a cabo su cálculo, la función más utilizada incluye dos elementos: la relación inversa existente entre la frecuencia absoluta de un término en la base de datos y su especificidad complementada con la frecuencia del término dentro del documento (en este caso con valoración positiva). Sin embargo, esta función sólo puede utilizarse para el caso del lenguaje natural, ya que para indización con lenguaje controlado la frecuencia de un término en cada documento siempre será igual a uno. La solución que propone el prof. Moya para salvar este problema consiste en sustituir ese segundo componente de la función por dos variables: la frecuencia del término en la clase o cluster (con valoración positiva) y el número de clases o clusters en los que dicho término o materia aparece (con valoración negativa). Como es evidente, esta función exige que previamente se haya procedido a incluir las materias en sus correspondientes clases o clusters, para lo que el autor propone un método muy simple, consistente en la utilización de los números de CDU asignados a los documentos. Además, estos clusters de materias generados a partir de la CDU pueden ser utilizados tam-

bién para mejorar las opciones de browsing del sistema, tan necesarias en sistemas dirigidos a usuarios finales como los OPACs.

Por otro lado, el autor también ha llevado a cabo pruebas formales para comparar cuál de las cuatro funciones utilizadas habitualmente para calcular la similaridad (Dice, Jaccard, Coseno y producto escalar) es más adecuada en el contexto de los SIGBs. En su opinión, la función del producto escalar es la que mejor se adapta a dicho contorno. Evidentemente, tanto para alcanzar esta conclusión como para proponer la función de cálculo del valor de discriminación definida con anterioridad, el autor se ha apoyado en una intensa labor de experimentación, cuyos resultados aparecen representados gráficamente.

Finalmente, hay que apuntar en el debe de esta monografía la utilización excesiva de anglicismos, algunos de ellos claramente evitables: performance, query... También se echa de menos un índice analítico al final de la obra que habría ayudado a una mejor localización de su contenido.

## REFERENCIAS

1. JARVELIN, K.; VAKKARI, P. The evolution of library and information science 1965-1985: a content analysis of journal articles. *Information Processing and Management*, 29(1), 1993, 129-144.
2. PERSSON, O. The intellectual base and research fronts of JASIS 1986-1990. *Journal of the American Society for Information Science*, 45(1), 1994, 31-38.

J. CARLOS FERNÁNDEZ MOLINA

---

---

NORMAS. DOCUMENTACIÓN. *Normas Fundamentales. Recopilación de Normas UNE*. AENOR [Fernández de la Hoz, 52. 28010 Madrid]. Madrid: AENOR, 1994. 269 p. ISBN: 84-86688-91-4. Signatura CDU: 006.72:02 (460) UNE. Precio: 4.990.- pesetas.

El manual incluye los textos completos de 32 normas UNE referentes a vocabularios, bibliotecas, publicaciones y bibliografías, por tanto va destinado a bibliotecarios, documentalistas, editores y también investigadores. Son las consideradas como fundamentales o básicas.

Hasta ahora el acceso a las Normas de Documentación era muy exclusivo, sólo se publicaban los proyectos relativos a información y documentación durante el período de preceptiva información pública, la fase de PNE (Proyecto de Norma Española) a UNE (Una Norma Española), en la Revista Española de Documentación Científica, después AENOR las editaba una a una y los autores, profesores, etc., las difundían poco a poco. Hasta aquí todo normal, el proceso lógico, pero al menos que visitásemos la Asociación Española de Normalización y Certificación, AENOR, donde nos dan toda clase de facilidades, no podías saber cuál era el conjunto de Normas, unas 59 aprobadas para 1993, incluso cuáles otras estaban en trámite, unas siete en 1995 y cuáles de todas ellas eran las fundamentales. Ahora este libro resuelve este problema.

Considerando el precio unitario de una sola Norma y multiplicándolo por el total de las recopiladas aquí, es una publicación muy barata para comenzar tu personal biblioteca sobre el tema.

Esperamos que AENOR siga incorporando a la publicación que comento ediciones de otras Normas de Documentación y reediciones modernas, cuando pasen el proceso de tramitación, que antes he explicado, y también que incorporen algunas más, de las Normas de su catálogo, no recogidas aquí, o al menos que añadan una lista completa, incluidas las que están en tramitación o en información pública, con lo cual el usuario podía completar su colección según su criterio.

Asimismo el elenco del estudioso pudiera tener otros títulos interesantes, no necesariamente de Documentación, según sus objetivos.

Finalmente también han publicado una "Guía para uso de la CDU", una guía introductoria para el uso y aplicación de la CDU de 122 páginas, muy bien cuidada, con lo que siguen la tradición de la Clasificación Decimal Universal para la organización del conocimiento, en su versión española, y que ya está en estudio otra nueva edición. Así como el Catálogo general de Normas, próximo.

El libro puede adquirirse en AENOR, BELSA, LUA y otras librerías especializadas. Lo mismo para las Tablas CDU, en versiones papel y disquetes para ordenador.

