
Gestión digital de la información: de bits a bibliotecas digitales y la web

Peña, Rosalía; Baeza-Yates, Ricardo; Rodríguez, José V. *Gestión digital de la información: de bits a bibliotecas digitales y la web*. Madrid: Ra-Ma, 2002. Isbn: 84-7897-514-4.

Ha llegado a la redacción de EPI este libro de la editorial *Ra-Ma* publicado en mayo de 2002, escrito por **Rosalía Peña** (*Univ. de Alcalá de Henares*) en colaboración con **Ricardo Baeza-Yates** (*Univ. de Chile*) y **José V. Rodríguez** (*Univ. de Murcia*).

Lo primero que llama la atención al hojear el libro es el índice por el detalle al que llega, algo muy útil si se pretende hacer una lectura no secuencial, ya que permite localizar más fácilmente los apartados que contienen determinada información de interés para el lector.

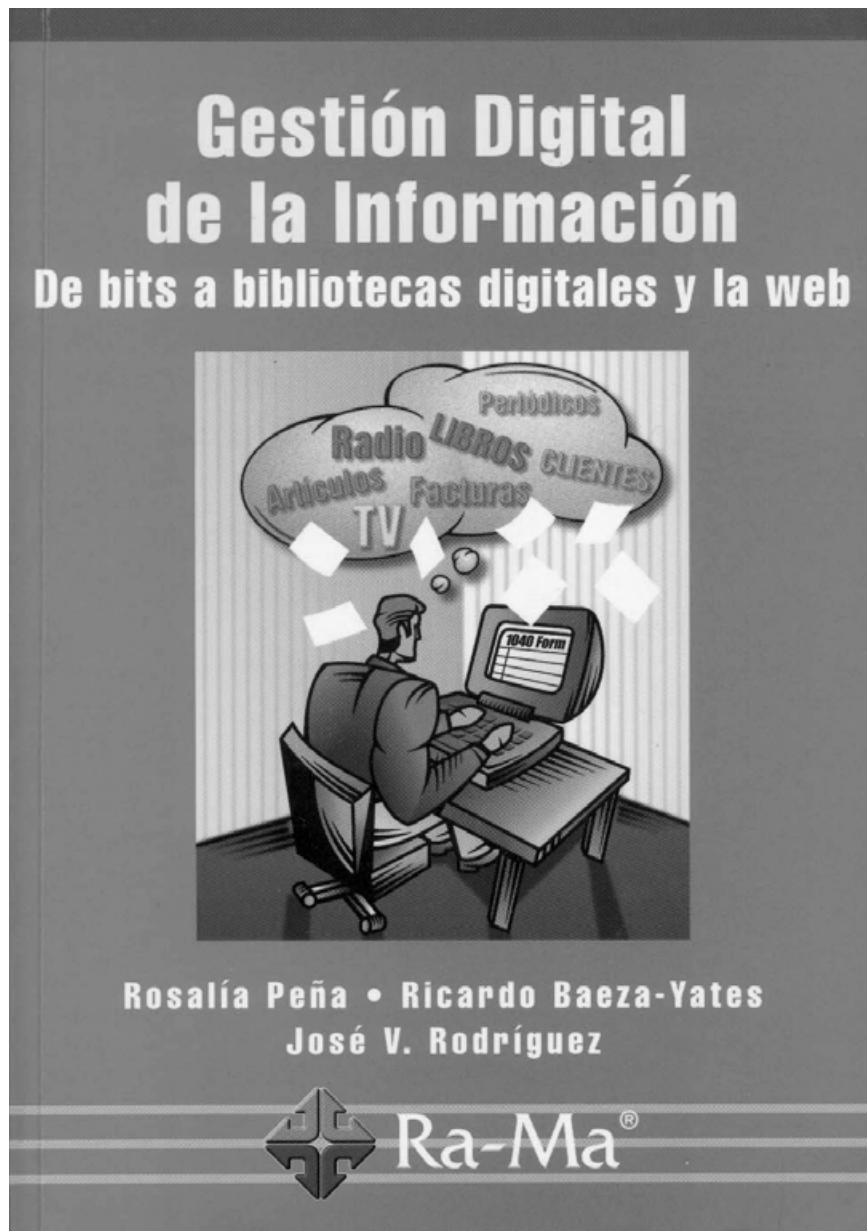
Su prólogo deja claro el contexto en el que se escribe la obra, los objetivos que pretende alcanzar, los lectores a los que se dirige



y la estructura que se le ha dado. En relación con el contexto, los autores justifican la necesidad de una obra cuya máxima aportación no es tanto de contenido como de sistematización de una materia hoy en día fundamental: la gestión de la información. Las últimas décadas han destacado por un aumento de la información que llega a desbordarnos, de ahí la creación de bases de datos, primero para gestionar datos numéricos y estructurados (los sgbd) y más adelante para tratar información más variable y desestructurada (los SRI y los posteriores buscadores de la web), pero siempre con el mismo objetivo: “contar con información pertinen-

te, precisa y en el momento oportuno”. Esa necesidad “ha provocado que los sistemas de recuperación de información estén en un momento crucial de cambio, ha generado el desarrollo de sistemas más avanzados que tratan de agregar información semántica a los documentos; de otra parte, la necesidad de acercar la herramienta al usuario desencadena una amplia actividad investigadora en torno a interfaces de usuarios, nuevos métodos de búsqueda, nuevas formas de indización y nuevos servicios”.

Con la obra, los autores se proponen “presentar los fundamentos del estado actual de los sistemas automatizados de acceso a la infor-



Parte I: Gestión de datos

Capítulo 1: Estructura de la información para su tratamiento

Capítulo 2: Estructuras complejas de datos

Capítulo 3: Soportes y métodos de acceso a ficheros

Capítulo 4: Organizaciones de ficheros

Capítulo 5: Bases de datos

Capítulo 6: Sistemas gestores de datos de gestión

Capítulo 7: Calidad de datos

Parte II: Recuperación de la información

Capítulo 8: Sistemas de recuperación de la información

Capítulo 9: Modelos booleanos

Capítulo 10: Modelo espacio vectorial

Capítulo 11: Perfeccionamiento de la recuperación

Capítulo 12: Evaluación de la recuperación

Capítulo 13: Inclusión de documentos en la base de datos

Capítulo 14: Búsqueda de información en la web

Capítulo 15: Bibliotecas digitales: integración de servicios, recursos y espacios informativos en redes

Anexo 1: Álgebra de Boole

Anexo 2: Xml

mación [...] y aproximar al lector no informático, de una manera sencilla e intuitiva, a las bases que le permitan saber qué puede o qué no puede esperar de un tratamiento automatizado y cómo estructurar la información para que pueda ser accedida de forma eficiente por el sistema”.

El libro se divide en dos partes bien diferenciadas, incluso en su orientación, que pueden leerse de forma independiente:

Gestión de datos

Da una visión general del tratamiento automatizado de datos y

trata la información desde el punto de vista informático. Son siete capítulos en los que se introducen los conceptos necesarios para entender la gestión de la información digital y se presentan distintas maneras de almacenarla en el ordenador. En ellos se trata la estructura y el almacenamiento de los datos en ficheros, los tipos de datos, la gestión de ficheros, los soportes de almacenamiento y acceso a la información, y la organización de los ficheros. Los capítulos 5 y 6 se dedican en las bases de datos como sistemas de estructura, acceso y gestión de la información, y el último de esta parte, el 7, se dedica a la calidad de los datos, que abarca la integridad, la restauración, la seguridad y la privacidad. Todo ello se presenta con un nivel de detalle suficiente para que un lector no familiarizado con la informática pueda comprender los conceptos más complejos sin perderse.

Recuperación de información

La segunda parte hace un completo recorrido por todos los aspectos de la recuperación automatizada de la información. Como se dice en el prólogo, esta parte “presenta el contexto de la recuperación de información y cómo estos sistemas tienen que lidiar con la ambigüedad del lenguaje natural, la diversidad y posible carencia de estructura de los documentos, surgiendo el concepto de pertinencia y otros parámetros de evaluación de los sistemas”. Los cinco primeros capítulos (del 8 al 12) explican qué son las bases de datos documentales (a las que llama “de recuperación”) y qué legislación les afecta (el derecho al acceso a la información frente al derecho a la propiedad intelectual). De una forma muy clara y didáctica se dedica un capítulo a los modelos booleanos y otro a los vectoriales de recuperación de información, aludiendo también a los modelos basados en

lógica difusa y a aquellos basados en el lenguaje natural. Se da importancia a la posibilidad del refinamiento de las consultas una vez ejecutada la búsqueda, de hecho el capítulo 11 trata de la reformulación de la consulta y la retroalimentación por relevancia. Este bloque finaliza con el capítulo 12, que está dedicado a evaluar los resultados obtenidos con criterios ya establecidos. Los tres últimos capítulos están algo más desvinculados de los anteriores: el 13 trata de la inclusión de documentos multimedia en las bases de datos, el 14, escrito por **Baeza-Yates**, se centra en la búsqueda de información en la web, y el último, de **Rodríguez Muñoz**, trata de las bibliotecas digitales.

«La primera parte introduce los conceptos necesarios para gestionar información digital, por lo que no es preciso que el lector tenga formación previa en informática»

El libro finaliza con un anexo del álgebra de Boole y otro sobre xml, además de una actualizada bibliografía y un completo índice de materias.

Por su estructura y los índices (el sumario y el índice de materias), el libro permite elegir una lectura secuencial o bien hipertextual, buscando los apartados que más interesan al lector en cada momento. Los conocimientos previos del lector y sus intereses determinarán la conveniencia de realizar un tipo u otro de orden en la lectura.

El lector profesional y el lector académico

Los objetivos de la obra se pretenden cumplir en dos tipos de lectores: los profesionales de la documentación por un lado, y el ámbito

¿Recibes ya EPI en casa?

Por sólo 69 euros + 4% IVA puedes tener todos los meses tu copia particular y leerla cómodamente cuando quieras.

académico universitario por otro, en este segundo caso se cubrirían a su vez dos ramas de estudio, la biblioteconomía y documentación, y la ingeniería informática. Los alumnos pueden comprobar la asimilación de los conceptos realizando los ejercicios que se proponen en cada capítulo y autoevaluándose en el último apartado de cada uno, titulado siempre “solución a ejercicios propuestos”. De la misma forma, los profesores encontrarán en esta obra parte de la materia que imparten presentada de una forma ordenada y sistematizada, con ejemplos clarificadores de conceptos a menudo difíciles de explicar a sus alumnos.

«Los sistemas de recuperación de información están en un momento crucial de cambio. Se están desarrollando nuevos métodos de búsqueda, sistemas que agregan información semántica a los documentos, nuevas formas de indización y nuevos servicios»

Este libro llega en un momento en que se hacía necesaria una recopilación en español sobre la gestión de la información, aunando temas que hasta ahora se han tratado más desde la informática y otros más propios de la documentación, pero que sin duda conviven y tendrán que hacerlo cada vez más. Tenerlo a mano no estará de más.

Reseña realizada por **Mari Carmen Marcos**
mcmarcos@posta.unizar.es