

## REVISIÓN DE LIBROS

# Modern Information Retrieval

**Ricardo Baeza-Yates y Berthier Ribeiro-Neto**

Addison Wesley

Harlow, England, 1999

Revisado por:

*José A. Pino, Universidad de Chile*

La aparición de este libro viene a llenar una sentida necesidad de investigadores y estudiantes de contar con un texto que incorporara avances en el área de Recuperación de información publicados en los últimos años en revistas científicas, conferencias y *workshops*. Asimismo, se requería un esfuerzo unificador de estos aportes, que muchas veces usan notaciones, terminología y enfoques muy dispares. Además, era necesario incorporar el fenómeno Internet como una motivación para la investigación y una realidad con la cual es necesario convivir hoy en día.

El libro de Baeza-Yates y Ribeiro-Neto hace eso y mucho más. Puede ser usado como texto básico de un curso de graduación o de postgrado, pero además, como una referencia importante en la actividad de investigación científica en el tema.

En lo que sigue, se comentan brevemente los capítulos de este texto.

El capítulo 1 (Introducción) presenta la perspectiva de los autores sobre el tema y sugerencias de cómo usar el texto en una variada gama de situaciones docentes.

El capítulo 2 (Modelamiento) es un aporte conceptual interesante al clasificar diversos modelos de Recuperación de Información coherentemente. Cada uno de ellos aparece presentado y al final hay secciones de "trends and research issues" y discusión bibliográfica.

El capítulo 3 (Retrieval Evaluation) se refiere a la evaluación de sistemas de recuperación de información. A pesar de su brevedad, el capítulo presenta, además de las medidas clásicas de evaluación, las colecciones de referencia (TREC, ACM, ISI).

El capítulo 4 (Lenguajes de Consulta), escrito conjuntamente con Gonzalo Navarro, presenta en forma compacta, los diversos lenguajes de consulta, en la que se agregan secciones de temas de investigación y discusión bibliográfica, muy útiles para los estudiantes de postgrado.

El capítulo 5 (Operaciones de Consulta) también es un valioso aporte al ordenar en forma muy coherente los diversos enfoques que se han seguido para mejorar formulaciones de consultas a través de su expansión y re-escritura de términos.

El capítulo 6 (Texto y Lenguajes Multimedia) revisa brevemente el lenguaje natural, los lenguajes de marcas y multimedia.

El capítulo 7 (Operaciones Sobre Texto), escrito con Nivio Ziviani, incluye pre-procesamiento de texto (temas de eliminación de palabra de ruido y el uso de *thesauri*, entre otros) y compresión de texto (codificación de Huffman, compresión de archivo invertido).

El capítulo 8 (Indexación y Búsqueda), escrito con Gonzalo Navarro, incluye los temas de Archivos Invertidos, Arreglos de sufijos, archivos de firmas, varios algoritmos de búsqueda secuencial y coincidencia de patrones, entre otros. Nuevamente, es útil la perspectiva de "trends and research issues" proporcionada por los autores, que son investigadores especializados en el tema.

El capítulo 9 (Recuperación de Información Paralela y Distribuida), escrito por Eric Brown, presenta de manera fluida y didáctica las principales ideas de procesamiento paralelo y distribuido aplicados a la recuperación de información.

El capítulo 10 (Interfaces de Usuario y Visualización), escrito por Marti Hearst, es una panorámica muy completa del tema Interacción Humano-Computador referido a Recuperación de información. Se incluye: el proceso de acceso a la información, la especificación de consultas, los contextos de visualización, el uso de juicios de relevancia y el soporte de la interfaz para el proceso de búsqueda.

El capítulo 11 (Recuperación de Información de Multimedia), escrito por Elisa Bertino, Bárbara Catania y Elena Ferrari, presenta el modelado y la especificación de consultas en el caso de información tipo multimedia.

El capítulo 12 (Indexación y Búsqueda en Recuperación de Información de Multimedia), escrito por Christos Faloutsos, presenta las ideas del enfoque GEMINI para tratar el problema de la indexación y búsqueda de información multimedia.

El capítulo 13 (Buscando en la Web) es otro de los aportes más novedosos de este libro, al incluir, en un solo texto coherente, un modelo de la Web, Máquinas de búsqueda, "browsing", "metasearches", búsqueda mediante "hyperlinks".

El capítulo 14 (Bibliotecas y Sistemas Bibliográficos), escrito por Edie Rasmussen, hace una revisión de sistemas de recuperación en línea y catálogos de acceso público en línea.

El capítulo 15 (Bibliotecas Digitales), escrito por Edward Fox y Ohm Sornil, presenta temas de organización de estas

bibliotecas, modelos de documentos, representaciones y acceso, y sobre los estándares en los que se basan.

Finalmente, el texto incluye un glosario y una importante bibliografía de 852 temas.

Es interesante destacar que el libro tiene una página web con cuatro lugares, a los cuales los lectores pueden hacer sugerencias, o contactar a los autores. Éstas son:

Brasil:

<http://www.dcc.ufmg.br/irbook>

Chile:

<http://sunsite.dcc.uchile.cl/irbook>

Italia:

<http://pluto.sm.dsi.unimi.it/irbook>

EEUU:

<http://www.sims.berkeley.edu/~hearst/irbook>

Este excelente libro de 513 páginas podría hacerse aun más valioso para la docencia si en una edición posterior incluyese ejercicios sugeridos para los estudiantes, al final de cada capítulo, los cuales se han prometido, lo que puede leerse en la página Web del libro.

