
RESEÑA

CACHEDA SEIJO, FIDEL; FERNÁNDEZ LUNA, JUAN MANUEL; HUETE GUADIX, JUAN FRANCISCO. *Recuperación de la información: un enfoque práctico y multidisciplinar*. Madrid: Ra-Ma, 2011, 811 págs. ISBN: 978-84-9964-112-6.

Recuperación de la información: un enfoque práctico y multidisciplinar ofrece al estudiante, al docente y al lector una visión general de la disciplina de la Recuperación de Información, abarcando desde las cuestiones fundamentales hasta los temas de investigación actuales.

La obra está coordinada por Fidel Cacheda Seijo, profesor del Departamento de Tecnologías de la Información y las Comunicaciones de la Universidad de A Coruña, Juan Manuel Fernández Luna y Juan Francisco Huete Guadix, profesores del Departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada.

Está concebido con un enfoque docente y formativo, respondiendo a la necesidad de disponer de un manual de referencia en español para las asignaturas relacionadas con la disciplina, impartidas en grados y másteres de Informática y Documentación. Ha contado con la participación de un grupo de investigadores y docentes expertos en las temáticas de los capítulos que desarrollan, la gran mayoría con experiencia docente en asignaturas relacionadas con la Recuperación de Información impartidas en las Facultades de Informática y en las Facultades de Documentación. La mayor parte de estos profesores e investigadores pertenecen a las áreas de Ciencias de la Computación y Lenguajes y Sistemas Informáticos, algunos de ellos relacionados con la informática aplicada a la Información y Documentación, como Enrique Herrera-Viedma,

Piedad Garrido, Ángel Zazo, José L. Alonso Berrocal, Carlos García-Figuerola, Fidel Cacheda, Pablo de la Fuente o Jesús M. Vegas. Para completar el enfoque multidisciplinar, el punto de vista de la Información y Documentación se proporciona con la participación de varios profesores del área de Biblioteconomía y Documentación, como Lluís Codina, Juan Antonio Martínez Comeche, Eduardo Peis o Jesús Tramullas.

El libro estructura los capítulos en tres grandes módulos de complejidad creciente (inicial, intermedio y avanzado). El enfoque docente y práctico se puede observar en la estructuración de cada capítulo: un resumen de los contenidos, exposición de los objetivos de aprendizaje, un desarrollo expositivo apoyado en ejemplos sencillos y una relación de ejercicios propuestos, algunos de ellos orientados a lectores con conocimientos de programación. El libro cuenta con material adicional que se puede descargar desde la web del editor: bibliografía básica comentada, recursos complementarios y soluciones a algunos ejercicios propuestos.

En el módulo inicial se presentan los fundamentos de la disciplina.

El capítulo 1 caracteriza la disciplina, define los principales conceptos y realiza una revisión de su historia y de las áreas que la componen. El capítulo 2 se centra en el procesamiento de documentos y consultas y en la construcción de ficheros invertidos. El capítulo 3 presenta los modelos clásicos de recupe-

ración (booleano, vectorial y probabilístico) y las técnicas de ponderación. El capítulo trata sobre las metodologías y medidas de evaluación de la recuperación. El capítulo 5 muestra las particularidades y evaluación de la recuperación de la información en la web. El capítulo 6 (Sistemas de búsqueda y obtención de información) estudia las características de los sistemas de información documental reales existentes y la formulación de estrategias de búsqueda y obtención de información (information seeking) del usuario en la web, bases de datos textuales, bancos de datos de imágenes y video y en repositorios.

En el módulo intermedio se abordan aspectos más específicos y complementarios de los temas del módulo inicial.

El capítulo 7 y 8 pueden entenderse como una extensión del capítulo 2, centrándose en los motores de indización de código abierto (apache lucene,...) y en las técnicas de construcción y compresión de índices. El capítulo 9 complementa al capítulo 3 presentando variantes de los modelos de recuperación clásicos como la indización por semántica latente. El capítulo 10 trata aspectos de los sistemas de recuperación de información relacionados con la reformulación de consultas, desarrollando las técnicas de realimentación por relevancia y de expansión de consultas, desarrollando las relacionadas con la utilización de tesauros. Los capítulos 11 y 12 ofrecen una visión general de la clasificación y el agrupamiento documental, mostrando las técnicas de aprendizaje supervisado y no supervisado en el dominio de las colecciones documentales. El capítulo 13 se centra en la estructuración de información, indexación, consulta y recuperación de documentos en XML. El capítulo 14 puede entenderse como un complemento

del capítulo 6 al tratar aspectos relacionados con el comportamiento en la búsqueda y uso de información de los usuarios.

El capítulo 15 introduce los sistemas de recomendación, desarrollando los aspectos relacionados con el filtrado de información colaborativo y los sistemas de recomendación basados en el contenido. El capítulo 16 se centra en un ámbito de aplicación de la recuperación como las bibliotecas digitales, desarrollando aspectos relacionados con su propósito, la representación (metadatos) y almacenamiento de objetos digitales, la interoperabilidad (protocolo OAI-PMH) y los modelos de bibliotecas digitales.

Por último, el tercer módulo, formado por los capítulos del 17 al 23, revisa varias técnicas más avanzadas y novedosas, en algunos aspectos relacionadas con propuestas de investigación. Comienza por el capítulo 17 que trata sobre el procesamiento de imágenes, la extracción y representación de propiedades de imágenes, y la recuperación de imágenes mediante consultas basadas en el contenido. El capítulo 18 analiza los aspectos avanzados de la implementación de motores de búsqueda como el crawling distribuido o los índices distribuidos. El capítulo 19 describe las principales técnicas de extracción de conocimiento (minería de datos) aplicadas al contenido, estructura y uso de la web. El capítulo 20 introduce el concepto de web semántica y se centra en la descripción de recursos y la recuperación de información mediante tecnologías de la web semánticas (RDF, SPARQL). Los capítulos 21, 22 y 23 son los recomendados para la configuración de temas de investigación en recuperación de información en doctorados, revisando varias técnicas avanzadas de recuperación de

información como el procesamiento del lenguaje natural y su relación con la recuperación de información, los sistemas de búsqueda de respuestas, la detección y seguimiento de temas de actualidad, la detección de novedades,...

Damos pues nuestro reconocimiento al esfuerzo de los autores para proporcionarnos una obra que seguramente será de utilidad para informáticos y documentalistas. La claridad expositiva de los capítulos y la sencillez de los

ejemplos propuestos pueden facilitar el aprendizaje a los estudiantes. Por otra parte, los docentes pueden encontrar en este manual un texto base sobre el que estructurar las asignaturas mediante las orientaciones, adaptaciones y ampliaciones de los diferentes temas propuestos.

Pedro Manuel Díaz Ortuño
Universidad de Murcia