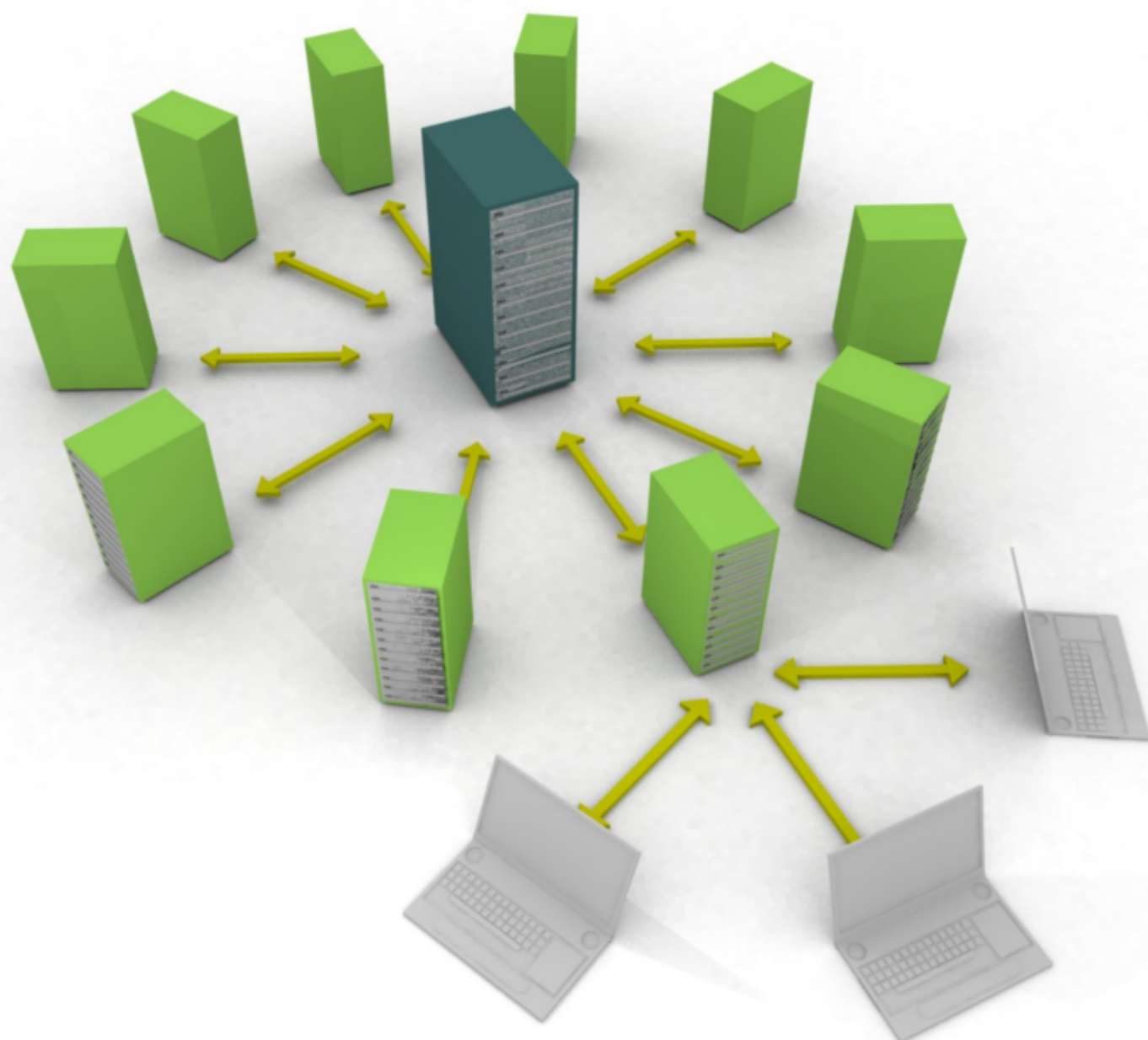


Tópicos avanzados de Bases de datos



AUTORES

Cristina Marta Bender

Claudia Deco

Juan Sebastián González Sanabria

María Hallo

Julio César Ponce Gallegos

Tópicos Avanzados de Bases de Datos

1a ed. - Iniciativa Latinoamericana de Libros de Texto Abiertos (LATIn), 2014. 115 pag.

Primera Edición: Marzo 2014

Iniciativa Latinoamericana de Libros de Texto Abiertos (LATIn)

<http://www.proyectolatin.org/>



Los textos de este libro se distribuyen bajo una licencia Reconocimiento-CompartirIgual 3.0 Unported (CC BY-SA 3.0) http://creativecommons.org/licenses/by-sa/3.0/deed.es_ES

Esta licencia permite:

Compartir: copiar y redistribuir el material en cualquier medio o formato.

Adaptar: remezclar, transformar y crear a partir del material para cualquier finalidad.

Siempre que se cumplan las siguientes condiciones:



Reconocimiento. Debe reconocer adecuadamente la autoría, proporcionar un enlace a la licencia e indicar si se han realizado cambios. Puede hacerlo de cualquier manera razonable, pero no de una manera que sugiera que tiene el apoyo del licenciador o lo recibe por el uso que hace.



CompartirIgual — Si remezcla, transforma o crea a partir del material, deberá difundir sus contribuciones bajo **la misma licencia que el original**.

Las figuras e ilustraciones que aparecen en el libro son de autoría de los respectivos autores. De aquellas figuras o ilustraciones que no son realizadas por los autores, se coloca la referencia respectiva.



Este texto forma parte de la Iniciativa Latinoamericana de Libros de Texto abiertos (LATIn), proyecto financiado por la Unión Europea en el marco de su Programa ALFA III EuropeAid.

El Proyecto LATIn está conformado por: Escuela Superior Politécnica del Litoral, Ecuador (ESPOL); Universidad Autónoma de Aguascalientes, México (UAA), Universidad Católica de San Pablo, Perú (UCSP); Universidade Presbiteriana Mackenzie, Brasil (UPM); Universidad de la República, Uruguay (UdelaR); Universidad Nacional de Rosario, Argentina (UR); Universidad Central de Venezuela, Venezuela (UCV), Universidad Austral de Chile, Chile (UACH), Universidad del Cauca, Colombia (UNICAUCA), Katholieke Universiteit Leuven, Bélgica (KUL), Universidad de Alcalá, España (UAH), Université Paul Sabatier, Francia (UPS).

Índice general

| | | |
|--------------|--|----|
| Introducción | | 9 |
| 1 | Diseño de Bases de Datos | 13 |
| 1.1 | Introducción | 13 |
| 1.2 | Modelado Entidad-Relación (ER) | 13 |
| 1.2.1 | Elementos del Modelo ER | 14 |
| 1.3 | Ciclo de Vida de las Bases de datos | 22 |
| 1.3.1 | Diseño Conceptual | 23 |
| 1.3.2 | Diseño Lógico | 23 |
| 1.3.3 | Diseño Físico | 25 |
| 1.4 | Estándares de Diseño | 26 |
| 1.5 | Bibliografía | 27 |
| 2 | Sistemas de Recuperación de Información (SRI) | 29 |
| 2.1 | Introducción | 29 |
| 2.1.1 | Recuperación de Información | 29 |
| 2.1.2 | Recuperación de Información versus Recuperación de Datos | 30 |
| 2.1.3 | Componentes de un Sistema de Recuperación de Información | 30 |
| 2.1.4 | Areas relacionadas con la Recuperación de Información | 31 |
| 2.1.5 | Recuperación de Información en la Web | 32 |
| 2.2 | Modelos de Recuperación de información | 32 |
| 2.2.1 | Modelo Booleano | 33 |
| 2.2.2 | Modelo Espacio Vectorial | 34 |
| 2.2.3 | Modelo Probabilístico | 35 |
| 2.2.4 | Otros modelos | 36 |
| 2.3 | Lenguajes de consulta | 36 |
| 2.3.1 | Consultas basadas en <i>keywords</i> | 37 |
| 2.3.2 | Consultas contextuales | 37 |
| 2.3.3 | Consultas booleanas | 38 |
| 2.3.4 | Consultas en lenguaje natural | 38 |
| 2.3.5 | Otros tipos de consultas | 38 |
| 2.4 | Evaluación de la Recuperación de Información | 39 |
| 2.4.1 | Relevancia | 39 |
| 2.4.2 | Indicadores de la Recuperación de Información | 40 |

| | | |
|------------|---|-----------|
| 2.4.3 | Otros indicadores | 42 |
| 2.4.4 | Colecciones de referencia | 42 |
| 2.5 | Estrategia de búsqueda | 43 |
| 2.5.1 | Expansión semántica de la consulta | 43 |
| 2.5.2 | Recursos lingüísticos para la expansión | 44 |
| 2.5.3 | Caso de estudio | 47 |
| 2.6 | Indexado y búsqueda | 48 |
| 2.6.1 | Preprocesamiento de los documentos | 49 |
| 2.6.2 | Archivo Invertido | 49 |
| 2.6.3 | Ventajas y desventajas del Archivo Invertido | 51 |
| 2.6.4 | En la web | 51 |
| 2.7 | Algunas aplicaciones | 52 |
| 2.7.1 | Consultas en un sitio web | 52 |
| 2.7.2 | Búsquedas Inteligentes, Sistemas Recomendadores | 53 |
| 2.7.3 | Recuperación de información en la web | 55 |
| 2.7.4 | Búsqueda de información multilingüe | 55 |
| 2.7.5 | Otros Problemas para Pensar | 56 |
| 2.8 | Bibliografía | 57 |
| 3 | Bases de Datos y Web | 59 |
| 3.1 | Introducción | 59 |
| 3.2 | Datos Semiestructurados | 59 |
| 3.3 | XML (<i>eXtensive Markup Language</i>) | 60 |
| 3.4 | RDF (<i>Resource Description Framework</i>) | 62 |
| 3.5 | Ontologías y OWL (<i>Ontology Web Language</i>) | 63 |
| 3.6 | La Web Semántica | 65 |
| 3.7 | Bibliografía | 66 |
| 4 | Sistemas de ayuda a la toma de decisión | 67 |
| 4.1 | Introducción | 67 |
| 4.2 | Almacén de Datos (<i>Data Warehouse</i>) | 68 |
| 4.2.1 | Preparación del almacén de datos | 69 |
| 4.2.2 | Preparación de los datos | 70 |
| 4.3 | Modelo de datos | 70 |
| 4.4 | Procesamiento Analítico en Línea (OLAP) | 73 |
| 4.5 | Problemas y temas pendientes | 74 |
| 4.6 | Bibliografía | 74 |
| 5 | Minería de Datos | 75 |
| 5.1 | INTRODUCCIÓN | 75 |
| 5.2 | PREPROCESAMIENTO DE LOS DATOS | 77 |

| | | |
|-------------|---|------------|
| 5.3 | OLAP-OLTP | 80 |
| 5.4 | Arquitectura de un Sistema Típico de Minería de Datos | 81 |
| 5.4.1 | Arquitectura Centralizada | 82 |
| 5.4.2 | Arquitectura De Bases de Datos Cliente/Servidor | 82 |
| 5.5 | Otras Técnicas Computacionales en Minería de Datos | 83 |
| 5.5.1 | Sistemas Expertos | 84 |
| 5.5.2 | Lógica Difusa | 84 |
| 5.6 | Aplicaciones de la Minería de Datos | 84 |
| 5.7 | Herramientas Informáticas | 84 |
| 5.8 | ACTIVIDADES DE APRENDIZAJE | 85 |
| 5.9 | MATERIAL DE REFERENCIAS A CONSULTAR * OPCIONAL | 88 |
| 5.10 | Bibliografía | 88 |
| 6 | Datos Temporales y Datos Espaciales | 89 |
| 6.1 | Bases de Datos Temporales (<i>Temporal DataBases</i>) | 89 |
| 6.1.1 | Introducción | 89 |
| 6.1.2 | El Tiempo en las Bases de Datos Temporales | 90 |
| 6.1.3 | Extensiones al Modelo Relacional | 91 |
| 6.1.4 | El modelo relacional temporal y TSQL | 91 |
| 6.1.5 | Sistemas de Gestión de Bases de Datos Temporales | 91 |
| 6.1.6 | TSQL2 | 92 |
| 6.2 | Bases de Datos Espaciales (<i>Spatial DataBases</i>) | 93 |
| 6.2.1 | Introducción | 93 |
| 6.2.2 | Sistemas de gestión de Bases de Datos Espaciales | 93 |
| 6.2.3 | Operaciones espaciales | 95 |
| 6.2.4 | Formatos de datos espaciales | 96 |
| 6.2.5 | Extensiones espaciales en SQL. El caso de MySQL | 96 |
| 6.3 | Bases de Datos Espacio Temporales (<i>Spatio-Temporal Databases</i>) | 97 |
| 6.3.1 | Introducción | 97 |
| 6.3.2 | Consultas sobre Datos Espacio Temporales | 98 |
| 6.3.3 | Extensión Espacio-Temporal del Modelo Entidad-Relación | 98 |
| 6.3.4 | SQLST | 99 |
| 6.4 | Bibliografía | 101 |
| 6.4.1 | Bases de Datos Temporales | 101 |
| 6.4.2 | Bases de Datos Espaciales | 102 |
| 6.4.3 | Bases de Datos Espacio Temporales | 102 |
| 7 | Sistemas de Bases de datos NoSQL | 103 |
| 7.1 | Introducción | 103 |
| 7.2 | Sistemas de Administración de bases de datos NoSQL | 104 |
| 7.2.1 | Almacenamientos clave-valor | 104 |
| 7.2.2 | Almacenamientos de documentos | 106 |
| 7.2.3 | Sistema de Base de Datos de Grafos | 108 |

| | | |
|------------|--|------------|
| 7.3 | Tecnologías de Linked Data y repositorios RDF | 109 |
| 7.3.1 | Sistemas de gestión de repositorios RDF | 110 |
| 7.3.2 | Tecnologías relacionadas con <i>Linked Data</i> que interactúan con repositorios RDF. 110 | |
| 7.3.3 | Respositorios RDF basados en tecnologías NOSQL | 112 |
| 7.4 | Bibliografía | 113 |

Introducción

El objetivo de este libro es presentar a los estudiantes una introducción a las nuevas tendencias en bases de datos soportadas por la evolución de la tecnología informática, que contemplan la gestión de nuevos tipos de datos.

La propuesta está dirigida a segundas asignaturas de Bases de Datos y cubre temas tales como la gestión de bases de datos no SQL, datos temporales, datos espaciales, datos en la web, la recuperación de información documental, la creación de repositorios para ayudar a la toma de decisiones, incluyendo una introducción a técnicas y herramientas para analizar estos datos.

Los capítulos “Sistemas de Recuperación de Información”, “Bases de Datos y Web”, “Datos Temporales y Datos Espaciales”, y “Sistemas de ayuda a la toma de decisión”, son aportes de Cristina Bender y Claudia Deco. Juan Sebastián González Sanabria es autor del capítulo “Diseño de Bases de Datos”. El capítulo “Bases de Datos NoSQL” fue escrito por María Hallo. El capítulo “Minería de Datos” es una contribución de Julio César Ponce Gallegos.

Breve CV de los autores

Cristina Marta Bender

Universidad Nacional de Rosario, Facultad de Ciencias Exactas, Ingeniería y Agrimensura
Universidad Católica Argentina, Campus Rosario, Facultad de Química e Ingeniería
Argentina

Magister en Informática por la Universidad de la República, Uruguay; Ingeniera Electrónica por la Universidad Nacional de Rosario, Argentina. Profesora e investigadora en el Departamento de Sistemas e Informática de la Facultad de Ciencias Exactas, Ingeniería y Agrimensura, Universidad Nacional de Rosario; y en la Facultad Católica de Química e Ingeniería, Campus Rosario de la Universidad Católica Argentina. Participa en proyectos nacionales e internacionales de investigación vinculados a sus intereses de investigación. También es coautora de varios artículos en journals y proceedings de conferencias internacionales y nacionales. Además, es revisora de artículos para diversos congresos. En su actividad docente es tutora y evaluadora de tesinas de grado y de posgrado. Sus intereses de investigación incluyen tecnologías de bases de datos y de recuperación de información (information retrieval).

Claudia Deco

Universidad Nacional de Rosario, Facultad de Ciencias Exactas, Ingeniería y Agrimensura
Universidad Católica Argentina, Campus Rosario, Facultad de Química e Ingeniería
Argentina

Doctora en Ingeniería por la Universidad Nacional de Rosario, Argentina. Magister en Informáti-