Preface

Capturing and utilizing data is a top priority of most businesses. Databases are fundamental to both internal and external business operations, but they can prove to be an expensive proposition. Increasingly important data are tied to information stored or gained from a company’s Web site. Understandably, businesses throughout the world are exploring new technologies to optimize their database and Web site performance. For many organizations, data warehousing and data mining techniques offer promising improvements for the use and capture of data. Additionally, Web-base engineering is a significant development in the process of Web page design and development. In order to get the most from these emerging technologies, business people, academics, researchers and students alike need to have access to the most up-to-date information about the development and implementations of these technologies. This timely new book addresses the important issues of justifying the cost of data warehousing techniques, data allocation, data model development and successful use of Web-engineering technologies. The authors represent a wide variety of perspectives and provide insights from many different cultures and organizational and industrial backgrounds.

Chapter 1 entitled, “An Introduction to Information Technology and Business Intelligence” by Stephan Kudyba (USA) and Richard Hoptroff (The Netherlands) provides a comprehensive introduction to the use of information technology within the concept of business intelligence. The chapter looks at the technologies currently in use, describes current economic theory and applies them both to developing effective business strategies.

Chapter 2 entitled, “Some Issues in Design of Data Warehousing Systems” by Ladjel Bellatreche and Kamalakar Karlapalem, the University of Science and Technology (Hong Kong), and Mukesh Mohania, Western Michigan University (USA), defines data warehousing and treats two major problems in data warehousing, namely data partitioning and the interaction between indexes and materialized views. The chapter discusses these problems in depth and provides practical solutions to solving these problems. Additionally, the authors discuss the impact of these problems on future strategies for data warehousing.
Chapter 3 entitled “Benchmarking Data Mining Algorithms” by Balaji Rajagopalan of Oakland University and Ravi Krovi of the University of Akron (USA) reports on a study which sought to test the theory that machine-learning algorithms, which are under no assumptions, should outperform their traditional counterparts when mining business’ databases. The results presented in the chapter can be used as prescriptive guidelines for the applicability of data mining techniques.

Chapter 4 entitled, “Justifying Data Warehousing Investments” by Ram Kumar of the University of North Carolina-Charlotte (USA) outlines an approach to justifying data warehousing investments that is based on the concept of options in finance. The approach described is being increasingly recognized as superior to the traditional methods used by finance professionals.

Chapter 5 entitled, “Data Mining: A New Arsenal for Strategic Decision-Making” by Sufi Nazem and Bongsik Shin of the University of Nebraska-Omaha (USA) provides an overview of this emerging technology and related trends in the application of data mining. Because data mining activities are often confidential and highly proprietary in nature, they rarely enjoy an open discussion. This chapter addresses the issues associated with data mining from an organizational perspective.

Chapter 6 entitled, “What’s in a Name? Exploring the Metaphorical Implications of Data Warehousing in Concept and Practice” by Elizabeth Davidson of the University of Hawaii-Manoa (USA) examines the metaphorical implications of the data warehousing concept and presents the findings of an empirical study of a data warehousing project that illustrated limitations of metaphor in practice. The authors discuss the implications of this metaphorical analysis for both theory and practice.

Chapter 7 entitled, “Incremental Data Allocation and Reallocation in Distributed Database Systems” by Amita Goyal Chin of Virginia Commonwealth University (USA) presents the Partial REALLOCATE and Full REALLOCATE heuristics for efficient data reallocation. By allowing only incremental introduction of servers into the distributed database system, complexity is controlled and costs are minimized. The authors use simple examples and a simulator to provide a framework for data reallocation in distributed database systems. The framework discussed produces nearly optimal solutions when compared with exhaustive methods.

Chapter 8 entitled, “Using Business Rules Within a Design Process of Active Databases” by Youssef Amghar, Madjid Meziane and Andre Flory of the National Institute of Applied Sciences (France) proposes a uniform approach to modeling business rules such as active rules, integrity constraints, etc. The authors then extend the state diagrams that are widely used for dynamic modeling. Additionally, the authors outline new functionalities of Computer Aided Software Engineering to take into consideration the active database specificities.
Chapter 9 entitled, “A Methodology for Datawarehouse Design: Conceptual Modeling” by Jose Maria Cavero and Esperanza Marcos of Universidad Rey Juan Carlos, Mario Piattini of Universidad de Castilla-La Mancha, and Adolfo Sanchez of Cronos Iberica (Spain) presents a multidimensional data warehouse development methodology based on and integrated with a Public software development methodology. This chapter is written in response to the need for a generally accepted complete methodology for data warehouse design.

Chapter 10 entitled “Assessing and Improving the Quality of Knowledge Discovery Data” of Herna Viktor and Niek du Plooy of the University of Pretoria (South Africa) discusses the use of data mining and data generation techniques including feature selection, case selection and outlier detection, to assess and improve the quality of data. In this approach, redundant, low quality data are removed from the data repository, and high quality data patterns are dynamically added to the set. The chapter also looks at the relationship between data capturing and the social aspects of office work.

Chapter 11 entitled, “Complementing the Data Warehouse with Information Filtered from the Web” by Witold Abramowicz, Pawel Jan Kaleczynski and Krzysztof Wecel of The Poznan University of Economics (Poland) examines the requirements for profiling in the data warehousing environments. The authors explore many issues concerning personalization, information overflow, user models and situatedness. The authors then analyze the contributing factors of the filtering process and offer some points to be considered during the extension of the evaluated system.

Chapter 12 entitled, “Justification of Data Warehousing Projects” by Reinhard Jung and Robert Winter of the University of St. Gallen (Switzerland) analyzes the economic justification for data warehousing projects and presents the results of a large academia-industry collaboration in the area of the non-technical issues of data warehousing. The authors derive basic steps and responsibilities for the justification of data warehousing projects based on an analysis of traditional approaches to economic IT project justification.

Chapter 13 entitled, “A Survey of Spatial Data Mining Methods Databases and Statistics Point of View” by Karine Zeitouni of the University of Versailles (France) reviews the data mining methods that are combined with Geographic Information Systems (GIS) for carrying out spatial analysis of geographic data. The chapter looks at data mining functions applied to spatial analysis data and highlights their specificity compared with their application to classical data. The authors then discuss two current methods of implementing these data mining techniques and discuss the similarities and differences between the approaches.

Chapter 14 entitled, “Efficient Query Processing with Structural Join Indexing in an Object Relational Data Warehousing Environment” by Vivekanand
Gopalkrishnan and Qing Li of City University of Hong Kong and Kamalakar Karlapalem of University of Science and Technology, Hong Kong (China) demonstrates the efficacy of building semantic-rich hybrid data indexes incorporating Structural Join Index Hierarch (SJIH) and Object Relational Data Warehousing (ORDW) views. The authors demonstrate this technique by using a set of queries to use a hill-climbing heuristic algorithm to select optima SJIHs, thereby embedding query semantics into the indexing framework. The authors also analyze the effectiveness of their approach when compared to a pointer chasing approach.

Chapter 15 entitled, “An Electronic Commerce Framework for Small and Medium Enterprises” by Anne Banks Pidduck of the University of Waterloo-Ontario (Canada) and Quang Ngoc Tran of TurboLinux describes an electronic commerce framework for small businesses. The authors discuss various services that a typical small business may want to provide for its customers and offer possible technologies to implement the services. Finally, the authors propose a prototype to generate such a model.

Chapter 16 entitled, “The VLEG Based Production and Maintenance Process for Web-Based Learning Applications” by Jorg Schelhase and Udo Winand of Universitat GH Kassel, Fachbereich Wirtschaftswissenschaften (Germany) presents an application for the realization, management and maintenance of Web-based learning applications. The chapter looks specifically at the Web-engineering principles necessary for improving Web-based learning.

Chapter 17 entitled, “Specification of Components Based on the WebComposition Component Model” by Martin Gaedke of the University of Karlsruhe and Klaus Turowski of the University of Federal Armed Forces (Germany) illustrates the benefits of the WebComposition Component Model in overcoming the gap between implementation and design models. This gap has proven to be quite a difficulty in the use of modern software engineering practices applied to Web-Engineering. The authors further illustrate the usefulness of their model by applying it to real world-applications.

Chapter 18 entitled, “The Development of Ordered SQL Packages to Support Data Warehousing” by Wilfred Ng of the Hong Kong University of Science and Technology (China) and Mark Levene of the University of London (United Kingdom) propose the enhancement of database languages in order to manipulate user-defined data orderings. The chapter further extends the relational model to incorporate partial orderings into data domains and describes the ordered relational model. The authors then discuss the details of the generic operations arising from OSQL packages called OSQL_Time, OSQL_INCOMP and OSQL_FUZZY.

Data warehousing and data mining techniques offer much to businesses. These important techniques can be used to optimize databases and get the most
from a business’s data collecting efforts. Additionally, these chapters discuss the emerging technology of Web-based engineering and discuss the practicalities of implementing Web-based engineering as a beneficial alternative to traditional software engineering practices. The chapters in this book address the issues important to businesses implementing these techniques, researchers investigating them, and students studying them. From how to assess the quality of the data and to justify the expense of these techniques to comparisons of various Web-based engineering techniques and specific languages and methods used to implement them, the chapters contained herein provide insightful theoretical discussion as well as practical examples and case studies illustrating the concepts discussed. This book is a must-have for all those interested in understanding and applying the most up-to-date research and practice in data warehousing, data mining and Web-based engineering.

IRM Press
December 2001