Preface

Database technologies have a rich history of relevant developments immersed in a continuous evolution and consolidation process. Even more, during the last decades, they have evolved in such a way that almost all main software applications and modern information systems have a database as a core component. The information stored is usually accessed and manipulated by many application programs to perform business processes. In this sense, databases in any organization have provoked a profound impact and significant endeavors in their operability, and business assessments.

Moreover, data are one of the most valuable assets of any organization and the design of database applications is a factor of vital influence regarding the efficiency and manageability of their information systems. The extraordinary growth and widespread application of databases has reached a vast diversity of users with their own fields of development, their particular application requirements, and their own technological needs. In recent years, these facts have promoted the appearance of new interdisciplinary investigation areas. It is worthy of mention distributed real-time systems, data integration based on ontologies, collaborative software development, databases on the Web, spatio-temporal databases, multimedia databases, new scopes of database programming languages, and the appearance of new characteristics related to data quality, indexing and reengineering, among others.

A database management system (DBMS) contributes to these objectives by providing data persistence, efficient access and data integrity. By isolating the conceptual schema from the implementation schema, database systems guarantee data independence from storage techniques and offer Standard Query Language (SQL) which is the query language per excellence. In addition, by means of the management of users and their privileges, the DBMS can provide safe control access to data. While the control of concurrent access to data is managed through different protocols of transaction scheduling and varied locking techniques, backups and database recovery strategies allow database recovering after hardware or software failures. These capabilities—among others—have opened wide research fields exciting challenges, and major technological and conceptual changes in many features through their evolution.

The “Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends” provides a new and comprehensive knowledge compendium on databases. This handbook pulls together many relevant issues that researchers and practitioners have investigated, proposed or observed to solve diverse real-world problems with the help of databases, and provides a wide compilation of references to topics in the field of database systems and applications.

Since knowledge about databases and their entire environment has become an essential part of any education in computer science and the main subject of many research groups at universities and institutes all over the world, this handbook is an ideal source of knowledge for students, researchers, programmers, and database developers who may need speedy and reliable information, and authoritative references to current database areas, latest technologies and their practical applications. This handbook provides many articles offering coverage and definitions of the most important issues, basic concepts, trends, and
technologies in database field along with some papers presenting a theoretical foundation of relevant topics in such field.

This handbook is intended for a wide range of readers including computing students having basic knowledge on databases; teachers in charge of introductory and advanced courses on databases; researchers interested in specific areas related to their research, and practitioners facing database implementation choices. Curious and inexperienced readers will also find in this handbook many interesting articles, opening the gate to an invaluable knowledge about databases principles and novel applications. Experienced teachers will find a comprehensive compendium of teaching resources. The main endeavor of this handbook has been to grant access to essential core material for education, research and practice on database systems.

The handbook is composed by 93 articles from authoritative database researchers, focusing on object-oriented applications, multimedia data storing and management; also, on new fields of applications such as geospatial and temporal information systems, data warehousing and data mining, design methodologies, database languages and distributed databases, among other topics. Emerging areas that are becoming particularly mature are also faced. They include the integration of DBMSs into the World Wide Web; the effective support to the decision-making process in an organizational environment; the information visualization and the high performance database systems.

This “Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends” is a collaborative effort addressing the endeavors that raise the increasing need of improving the storage of information, the adaptation or adherence of conceptual modeling and design to newer paradigms, and the development of advanced applications related to the Internet, e-commerce, data warehousing and data mining. Leading specialists in each area; researchers with a vast experience on the topics covered by this volume; experts in the development of database systems in organizational environments; and teachers with accumulated experience teaching graduate and undergraduate courses have contributed with valuable chapters on their fields of expertise.

This handbook has been built as a compilation of papers with a quasi-standardized structure. Many articles may be included into more than one group, but the arrangement was made taking into account their main area. Interested readers will be able to compose their own groups by gathering articles which share keywords or term definitions. It differs from the typical databases books in that it offers a quite balanced treatment of the most relevant characteristics, languages and definitions of the core terms in each subject. Many articles offer an analytical approach so that the concepts presented can serve as the basis for the specification of future systems. Many articles have plenty of examples to show readers how to apply that material.

On the other hand, each article offers a profuse set of recommended references to current and settled literature on each topic. The “Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends” presents a sound grounding in the foundations of database technology and the state of the art; also, it covers other areas which are under exceptional development and spreading. Thus, the articles in this volume include a list with the key terms and concepts relevant to each topic along with their definitions. We have to thank our authors for the careful selection of terms they have made.

Section I: Conceptual Modeling

This first section groups a set of articles dealing with relevant topics related to conceptual modeling, current and traditional models, formal specifications, new paradigms and data warehousing. Among other subjects, this section includes original work on the entity relational model, completeness of the
information, capture of requirements for data warehousing, symbolic objects, temporary data, post-relational data models and data reengineering.

Sikha Bagui is the author of “Mapping Generalizations and Specializations and Categories to Relational Databases”. This paper discusses the implementation of generalizations and specializations in relational databases, along with the application of the concept of inheritance, in the context of the extended entity relationship (EER) model.

In “Bounded Cardinality and Symmetric Relationships”, Norman Pendegraft gives an overview on bounded cardinality and its links with symmetric relationships, highlighting some of the problems they present, and discussing their implementation in a relational database.

The article “A Paraconsistent Relational Data Model” by Navin Viswanath and Rajshekhar Sunderraman deals with the Closed World Assumption and the Open World Assumption. The first assumption is based on the completeness of the information stored in the database. Consequently, if a fact is not in the database, then its negation is true; under the second assumption, such negation should be explicitly stored to become true; otherwise nothing can be said about it. This article introduces a data model which is a generalization of the relational data model: “The Paraconsistent Relational Data Model”.

The paper “Managing Temporal Data” by Abdullah Uz Tansel reviews the issues in modeling and designing temporal databases based on the relational data model. Also, it addresses attribute time stamping and tuple time stamping techniques.

Richard C. Millham contributed with an article entitled “Data Reengineering of Legacy Systems”, which provides an overview on the transformation of legacy data from a sequential file system to a relational database, outlining the methods used in data reengineering to transform the program logic that access the database using wide spectrum language (WSL) as the intermediate representation of programs.

The three following articles have been authored by Elzbieta Malinowski. These contributions are tightly coupled as they deal with the MultiDim model which is a conceptual multidimensional model used for representing data requirements for data warehousing (DW) and on line analysis processing (OLAP) applications. The article “Different Kinds of Hierarchies in Multidimensional Models” describes the MultiDim model, showing its abilities to denote fact relationships, measures, dimensions, and hierarchies, for which novel classification is provided. Malinowski considers that DWs and OLAP can use relational storage, and presents how hierarchies can be mapped to the relational model. In “Spatial Data in Multidimensional Conceptual Models”, additional characteristics of the MultiDim conceptual model are explored. It is extended providing spatial support for different elements such as levels and other measures. These characteristics are explored in the context of a platform-independent conceptual model. Finally in “Requirement Specification and Conceptual Modeling for Data Warehouses”, Malinowski presents a proposal to cope with the lack of a methodological framework to guide developers through the different stages of the data warehouse design process. This proposal refers to the requirements specification and conceptual modeling phases for data warehouses design unifying already existing approaches by giving an overall perspective of the different alternatives available to designers.

In “Principles on Symbolic Data Analysis”, Héctor Oscar Nigro and Sandra Elizabeth González Cisaro revise the history, sources and fields of influence of Symbolic Data, providing formal definitions and semantics applied to such novel concept. They discuss how to handle null values, internal variations and rules using the symbolic data analysis approach which is based on the symbolic object model.

Six contributions written by different authors address the field of database evolution. Luiz Camolesi Júnior and Marina Teresa Pires Vieira coauthored the article “Database Engineering Supporting the Data Evolution”. Their contribution surveys on database evolution as a vast subject under constant discussion and innovation, focusing on the evolutionary process, and the different ways in which it can be approached. Regarding that schema evolution is a key research topic with an extensive literature built
up over the years, the article summarizes why database evolution itself becomes hard to manage, and
describes some proposed approaches to manage the evolution of a schema for a wide range of types of
databases. Another approach for the evolution of databases can be found in “Versioning Approach for
Database Evolution”, written by Hassina Bounif, who analyzes versioning-based courses of action taking
into account that the versioning principles can be applied universally to many different forms of data.
The next four articles are framed into two main approaches: Schema Evolution approach and Schema
Versioning approach. The following two are authored by Vincenzo Deufemia, Giuseppe Polese, and
Mario Vacca. The first article, “Evolutionary Database: State of the Art and Issues” the authors focus on
the recent introduction of evolutionary database methodologies which broaden the schema evolution and
versioning problems to a wider vision highlighting new and more challenging research problems. The
second one has been entitled “Interrogative Agents for Data Modeling” and examines the problem of
evolutionary data modeling process. The authors present the characteristics of this subject in the frame
of the agent paradigm, as the evolutionary data modeling can be seen as a process in active databases
able to change their beliefs and structure. Moreover, following the agent-oriented software engineering
(AOSE) view, the authors show that the use of tools and techniques from artificial intelligence (AI) can
help to face the problem of developing supporting tools to automate evolutionary data modeling. The
other two papers address schema evolution models in the context of data warehouses. “Schema Evolution
Models and Languages for Multidimensional Data Warehouses” coauthored by Edgard Benítez-Guer-
rero and Ericka-Janet Rechy-Ramírez provides a deep analysis of both approaches reviewing recent
research results on the subject, while the article “A Survey of Data Warehouse Model Evolution” written
by Cécile Favre, Fadila Bentayeb and Omar Boussaid compares both approaches using three different
criteria: functionality, deployment and performance.

“Document Versioning and XML in Digital Libraries” by M. Mercedes Martínez-González, is devoted
to digital libraries. The author analyzes the issues related to document versioning and main existing ap-
proaches, together with their pros and cons. Also, she discusses how digital libraries mirror the traditional
library and how they provide more services than those available in paper document libraries.

In the framework of the model-driven development (MDD) approach, Harith T. Al-Jumaily, Dolores
Cuadra and Paloma Martinez contributed with the article “MDD Approach for Maintaining Integrity
Constraints in Databases”. The authors analyze the semantic losses produced when logical elements are
not coincident with conceptual elements—with especial emphasis on multiplicity constraints—and how
to fix them, proposing a trigger system as the maintaining mechanism.

Pierre F. Tiako in his article entitled “Artifacts for Collaborative Software Development” provides
an overview on collaborative software development analyzing modeling processes, and also, environ-
ment artifacts involved.

Other contributions dealing with Conceptual Modeling aspects can be found in the section Logical
Modeling (Section II): “Horizontal Data Partitioning: Past, Present and Future” by Ladjel Bellatreche
and Database Reverse Engineering by Jean-Luc Hainaut, Jean Henrard, Didier Roland, Jean-Marc Hick
and Vincent Englebert and in the section Ontologies (Section V): “Ontologies Application to Knowledge
Discovery Process in Databases” by Héctor Oscar Nigro and Sandra Elizabeth González Cisaro and “On-
tology-Based Semantic Models for Databases”, by László Kovács, Péter Barabás and Tibor Répási.

Section II: Logical Modeling

For several decades, data modeling has been an aspect of the database world that has received many
contributions from researchers and also important feedback from practitioners. Subjects as data modeling
evolution, versioning, reverse engineering, and the impact of novel applications have driven research-
ers and practitioners to revisit well-established approaches to address the challenges such subjects are raising. This section contains valuable contributions focusing on such aspects.

In “Object-Relational Modeling”, Jaroslav Zendulka shows how an object-relational database schema can be modeled in Unified Modeling Language (UML). Firstly, the author clarifies the fact that UML contains no direct support: neither for capturing important features of relational databases nor for specific features of object-relational databases. Regarding the fact that such features are necessary for modeling data stored in a relational database and objects stored in an object-relational database at design levels subsequent to the conceptual one, the author describes an extension of UML which adds the ability to model effectively and intelligibly such features in this kind of databases.

“Concept-Oriented Model” by Alexandr Savinov reviews concept-oriented model (CoM), an original approach to data modeling he has recently introduced. Its major goal consists in providing simple and effective means for the representation and manipulation of multidimensional and hierarchical data while retaining the possibility to model the way data is physically represented.

From a tutorial perspective, in the article “Database Reverse Engineering”, Jean-Luc Hainaut, Jean Henrard, Didier Roland, Jean-Marc Hick and Vincent Englebert describe the problems that arise when trying to rebuild the documentation of a legacy database as long as the methods, techniques and tools that may be used to solve these problems.

“Imprecise Functional Dependencies” is a paper coauthored by Vincenzo Deufemia, Giuseppe Polese and Mario Vacca that overviews imprecise functional dependencies and provides a critical discussion of the dependencies applicable to fuzzy and multimedia data.

The article “Horizontal Data Partitioning: Past, Present and Future” by Ladjel Bellatreche is devoted to the analysis of the issues of horizontal data partition, the process of splitting access objects into sets of disjoint rows. This analysis ranges from the former utilizations—logically designing databases efficiently accessed—to the recent applications in the context of distributed environments and data warehouses.

Two contributions authored by Francisco A. C. Pinheiro are devoted to analyze the interaction of novel applications of database systems and the improvement of technologies and paradigms. The article “Database Support for Workflow Management Systems” provides an interesting overview on the relationships between database technologies and workflow issues. In this regard, the author addresses the discussion on how advances on databases as a supporting technology may be applied to build more useful workflow applications and how workflow application needs may drive the improvement of database technologies. On the other hand, the article “Politically Oriented Database Applications” deals with how technology pervades every aspect of modern life, having an impact on the democratic life of a nation and frequently, being an object of dispute and negotiation. These facts affect the way politics is done, by shaping new forms of planning and performing political actions. Applications used in or related to politics are information intensive, making databases a prime element in building politically oriented applications. In this article, Francisco A. C. Pinheiro discusses some aspects of database related technology necessary for this kind of applications.

Facing the need to integrate information efficiently, organizations have implemented enterprise resource planning (ERP) systems. Much of the value of these ERP systems resides in their integrated database and its associated data warehouse. Unfortunately, a significant portion of the value is lost if the database is not a semantic representation of the organization. Taking into account such negative aspect, Cheryl L. Dunn, Gregory J. Gerard, and Severin V. Grabski have coauthored the article “Semantically Modeled Databases in Integrated Enterprise Information Systems” focusing on the resources-events-agents (REA) ontology.

“The Linkcell Construct and Location-Aware Query Processing for Location-Referent Transactions in Mobile Business” contributed by James E. Wyse, describes location-qualified business information
for the provision of location-based mobile business. This information—contained in a locations repository—and its management—performed by a locations server—are the focal concerns of this article.

Hagen Höpfner is the author of “Caching, Hoarding, and Replication in Client/Server Information Systems with Mobile Clients”. This paper presents a complete set of exact definitions of the caching, hoarding and replication techniques for handling redundant data in information systems with mobile clients in relation to the level of autonomy of mobile devices/users. Furthermore, the author explains the terms cache replacement, cache invalidation, cache maintenance, automated hoarding, and synchronization of replicated data.

Section III: Spatial and Temporal Databases

Databases handling temporal, spatial or both types of data are becoming more and more frequently used every day. Temporal data contains some references, attributes, or structures where time plays a role; the same happens with spatial data. The integration of factual with temporal and/or spatial data to be able to handle geographical information systems (GIS), location based services, all kind of mapping services or weather services require a profound understanding of the special needs such integration demands. This section contains several specialized contributions related to these matters.

Two contributions written by the same authors, deal with the processing of spatial temporal databases. The first one, “Spatio-Temporal Indexing Techniques” by Michael Vassilakopoulos and Antonio Corral, surveys the indexing of moving points and other spatio-temporal information, considering recent research results and possible research trends within this area of raising importance. The second one, “Query Processing in Spatial Databases” by Antonio Corral and Michael Vassilakopoulos specifically focuses on spatial query processing.

Khaoula Mahmoudi and Sami Faïz in “Automatic Data Enrichment in GIS Through Condensate Textual Information” propose a modular approach to enrich data stored in a geographic database (GDB), by extracting knowledge from on-line textual documents corpora. This is accomplished by using a distributed multi-document summarization. A refinement step to improve the results of the summarization process based on thematic delimitation, theme identification, delegation and text filtering is also proposed.

From a tutorial perspective, Maria Kontaki, Apostolos N. Papadopoulos and Yannis Manolopoulos in their article “Similarity Search in Times Series” introduce the most important issues concerning similarity search in static and streaming time series databases, presenting fundamental concepts and techniques.

“Internet Map Services and Weather Data”, a contribution by Maurie Caitlin Kelly, Bernd J. Haupt and Ryan E. Baxter, provides a brief overview of the evolution and system architecture of internet map services (IMS), identifying some challenges related to the implementation of such service. The authors provide an example of how IMS have been developed using real-time weather data from the National Digital Forecast Database (NDFD).

The two following contributions address subjects related to spatial network databases. The first one, “Spatial Network Databases” by Michael Vassilakopoulos reviews the motivation behind the development of techniques for the management of spatial networks and their fundamental concepts. Additionally, the author reports the most representative and recent research efforts and discusses possible future research. The second one “Supporting Location-Based Services in Spatial Network Databases”, contributed by Xuegang Huang, summarizes existing efforts from the database community to support location-based services (LBSs) in spatial networks, focusing the discussion on the data models, data structures, and query processing techniques. The author considers a prototype service that finds the k nearest neighbors to a mobile user in the network.

Laura Díaz, Carlos Granell, and Michael Gould focus on the interoperability problem from a syntactic point of view. In their article “Spatial Data Integration Over the Web”, they propose the use of
interface standards as a key to spatial data integration over the Web. For that purpose, they report on the Geography Markup Language (GML) standard that provides spatial services with common data models for spatial data access and interchange of representation between spatial and non-spatial data with an XML-based format.

Other contributions dealing with Spatial and Temporal Databases aspects can be found in the section Conceptual Modeling (Section I): “Managing Temporal Data” by Abdullah Uz Tansel, in the section Ontologies (Section V): “Mediation and Ontology-Based Framework for Interoperability” by Leonid Stoimenov and in the section Physical Issues (Section VII): “Querical Data Networks” by Cyrus Shahabi and Farnoush Banaei-Kashani.

Section IV: Database Integrity

Database integrity is known to be important from the earliest days in the database history. At first glance, it could be said that database implementations have traded data redundancy or access flexibility to the data stored by new integrity requirements. When such flexibility reaches distributed databases or context aware applications, the integrity management needs to be increased again. It may be also true that future paradigms will raise new integrity issues. Several contributions related to integrity constraint checking, fault tolerant integrity control, and several points of views on data quality are included in this section.

In “Improving Constraints Checking in Distributed Databases with Complete, Sufficient, and Support Tests”, Ali Amer Alwan, Hamidah Ibrahim and Nur Izura Udzir analyze the performance of the checking process in a distributed environment when various types of integrity tests are considered. Authors select the most suitable test for each situation in terms of the amount of data transferred across the network and the number of sites involved during the process of checking the constraints.

The paper “Inconsistency-Tolerant Integrity Checking” by Hendrik Decker and Davide Martinenghi highlights the fact that integrity checking is practically unfeasible for significant amounts of stored data without a dedicated approach to optimize the process. The authors give a fresh perspective by showing that if the simplified form of an integrity theory is satisfied then, each instance of each constraint that has been satisfied in the old state continues to be satisfied in the updated state even if the old database is not fully consistent. They rightfully call this approach “inconsistency-tolerant”.

“Merging, Repairing, and Querying Inconsistent Databases”, the article contributed by Luciano Caroprese and Ester Zumpano, introduces a framework for merging, repairing and querying inconsistent databases, investigating the problem of the satisfaction of integrity constraints implemented and maintained in commercial DBMS in the presence of null values. The authors also establish a new semantics for constraints satisfaction.

“The Challenges of Checking Integrity Constraints in Centralized, Distributed and Parallel Databases” by Hamidah Ibrahim surveys on the vital problem of guaranteeing database consistency, highlighting several factors and issues as regards preventing semantic errors made by the users due to their carelessness or lack of knowledge.

The following two contributions examine quality of data issues. The first one “Data Quality Assessment” by Juliusz L. Kulikowski, tackles the basic problems of data quality assessment, assuming that for high information processing systems’ effectiveness high quality of data is a the main requirement. In “Measuring Data Quality in Context”, the authors Gunnesan Shankaranarayanan and Adir Even propose a framework to assess data quality within specific usage contexts linking it to data utility. The utility of data is conceived by these authors as a measure of the value associated with data within specific usage contexts.

Two related contributions sharing a couple of authors are devoted to data integrity in Geographical Information Systems. The first one, “Geometric Quality in Geographic Information” by José Francisco
Zelasco, Gaspar Porta and José Luis Fernandez Ausinaga proposes a method to evaluate the geometric
integrity of digital elevation models (DEM) obtained by different techniques. In the second one, “Geo-
metric Quality in Geographic Information IFSAR DEM Control”, José Francisco Zelasco, Judith Donayo,
Kevin Ennis and José Luis Fernandez Ausinaga consider Interferometry SAR (IFSAR) techniques and the
stochastic hypotheses that are specific according to the particular geometry involved in this technique.

“Querying and Integrating P2P Deductive Databases” contributed by Luciano Caroprese, Sergio
Greco and Ester Zumpano considers the integration of information and the computation of queries in
an open-ended network of distributed peers. This proposal is based on a change in the perception of
inconsistent peers, accepting data answering queries from those peers if it comes from the consistent
part of the peer.

Other contributions dealing with Database Integrity aspects can be found in the section Conceptual
Modeling (Section I): “Bounded Cardinality and Symmetric Relationships” by Norman Pendegraft,”
MDD Approach for Maintaining Integrity Constraints in Databases” by Harith T. Al-Jumaily, Dolores
Cuadra and Paloma Martínez and “A Paraconsistent Relational Data Model” by Navin Viswanath and
Rajeshwar Sunderraman and in the section Ontologies (Section V): “Inconsistency, Logic Databases
and Ontologies” co-authored by José A. Alonso-Jiménez, Joaquín Borrego-Díaz and Antonia M. Chávez-
González.

Section V: Ontologies

Interaction between the fields of ontologies and databases may be produced in several ways. Ontologies
may be used or required to understand the context of the future database applications being the founda-
tion of the database schema. A database may be used as a repository for large ontologies. Ontologies
may be also used to ease the integration of heterogeneous and distributed databases. This section holds
articles dealing with different interactions between ontology and databases.

The article “Using Semantic Web Tools for Ontologies Construction” by Gian Piero Zarri describes
the proper characteristics of the ontological approach that support the Semantic Web, differentiating it
from the ‘classical’ approach of the construction of ontologies based on a methodology of the ‘frame’
type and on the use of tools in the ‘standard’ Protégé style.

In “Matching Relational Schemata to Semantic Web Ontologies”, Polyxeni Katsiouli, Petros Pa-
papanagiotou, Vassileios Tsetsos, Christos Anagnostopoulos and Stathes Hadjieithymiades propose a
methodology for schema matching and present a tool called Ronto (Relational to ONTOlogy). This
tool deals with the semantic mapping between the elements of a relational schema to the elements of an
ontological schema, in the context of data migration.

“Ontology-Based Semantic Models for Databases” by László Kovács, Péter Barabás and Tibor
Répási shows and explains the importance and the role of ontologies in design processes, regarding the
recognition that ontologies have achieved as the description formalism for knowledge representation.

“Inconsistency, Logic Databases, and Ontologies” is an article co-authored by José A. Alonso-Ji-
ménez, Joaquín Borrego-Díaz, and Antonia M. Chávez-González. The authors base their contribution
on the fact that to work with very large databases makes certain techniques for inconsistency handling
not applicable. They discuss how in the semantic web future trends must study verification techniques
based on a sound and limited testing, aided by a powerful automated theorem prover. These techniques
need a deep analysis of the behavior of automated theorem provers having great autonomy because a
slanted behavior may produce deficient reports about inconsistencies in the knowledge database (KDB).
For these authors, the most promising research line in this field is the design and development of tools
that allow explaining the source of anomalies detected in ontologies.
The following four articles focus on the fact that geographical information systems (GIS) are increasingly moving away from monolithic systems towards the integration of heterogeneous and distributed information systems. This interoperability problem forces to deal with a diversity of data sets, data modeling concepts, data encoding techniques and storage structures. Furthermore, a problem of semantic heterogeneity arises: different data sets usually have discrepancy in the terms they use. Software systems do not have “common sense” –as humans do– to deal with these discrepancies. Software systems usually do not have any knowledge about the world, leading to serious conflicts while discovering and interpreting data. “Data Integration: Introducing Semantics” is a tutorial article contributed by Ismael Navas-Delgado and José F. Aldana-Montes in which the reader will find a simple description of the basic characteristics of the data integration systems and a review of the most important systems in this area (traditional and ontology-based), along with a table highlighting the differences between them. Two contributions written by the same authors also address the issues related to data integration from a tutorial point of view. Both articles are devoted to the use of ontologies in the integration process, noting the advantages they bring to such process. The first article “Overview of Ontology-Driven Data Integration” by Agustina Buccella and Alejandra Cechich deals with a wider perspective considering general purpose Database Systems, while in the second article “Current Approaches and Future Trends of Ontology-Driven Geographic Integration” the focus is on geographic data. Two main problems are addressed in this case. The first one is how to combine the geographical information available in two or more geographical information systems with different structures, and the second one is related to the differences in the points of views and vocabularies used in each geographical information system.

Leonid Stoimenov, author of “Mediation and Ontology-Based Framework for Interoperability” considers the interoperability problem in the context of geographical information systems (GIS). In his article, Stoimenov introduces a common interoperability kernel called ORHIDEA (ontology-resolved hybrid integration of data in e-applications) consisting of three key components: semantic mediators, translators/wrappers, and a shared server.

In “Ontologies Application to Knowledge Discovery Process in Databases”, the authors Héctor Oscar Nigro and Sandra Elizabeth González Cisaro discuss the application of ontologies in KDD, and propose a general ontology-based model, which includes all discovery steps.

Other contribution dealing with Ontologies aspects can be found in Physical Issues (Section VII): “Full-Text Manipulation in Databases”, by László Kovács and Domonkos Tikk.

Section VI: Data Mining

As data analysis techniques must process large amounts of data efficiently, special attention has been paid to new trends such as: evaluation techniques, safeguard of sensitive information and cluster techniques. Data mining is an interdisciplinary area of research, having its roots in databases, machine learning, and statistics. Several entries reporting many research efforts and main results in this field can be read in this section.

Edgard Benítez-Guerrero and Omar Nieva-García describe the problems involved in the design of an inductive query language and its associated evaluation techniques, and present some solutions to such problems in their article “Expression and Processing of Inductive Queries”. They also present a case study based on their proposal of an extension to SQL for extracting decision rules of the form if <conditions> then <class> to classify uncategorized data, and associated relational-like operators.

“Privacy Preserving Data Mining” (PPDM), an article contributed by Alexandre Evfimievski and Tyrone Grandison, reviews PPDM as the area of data mining that seeks safeguarding sensitive information from unsolicited or unsanctioned disclosure.
In “Mining Frequent Closed Itemsets for Association Rules”, Anamika Gupta, Shikha Gupta, and Naveen Kumar discuss the importance of mining frequent closed itemsets (FCI) instead of frequent itemsets (FI) in association rule discovery procedure, and explain different approaches and techniques for mining FCI in datasets.

“Similarity Retrieval and Cluster Analysis Using R*-Trees” contributed by Jiaxiong Pi, Yong Shi, and Zhengxin Chen examines time series data indexed through R*-Trees. The authors also study the issues of retrieval of data similar to a given query, and the clustering of the data based on similarity.

The paper entitled “Outlying Subspace Detection for High-dimensional Data” by Ji Zhang, Qigang Gao, and Hai Wang, gives an overview on the detection of objects that are considerably dissimilar, exceptional and inconsistent with respect to the majority of the records in an input database (outliers) and their outlying subspaces, i.e. subspaces in high-dimensional datasets in which they are embedded.

Two other contributions address data clustering issues. Clustering is one of the most important techniques in data mining. It is a tool to discover similar objects into different groups or non-overlapping clusters so that the data in each group shares commonality, often proximity, according to some defined distance measure. “Data Clustering” by Yanchang Zhao, Longbing Cao, Huaiifeng Zhang, and Chengqi Zhang provides a wider view on the clustering problem presenting a survey of popular approaches for data clustering, including well-known clustering techniques, such as partitioning clustering, hierarchical clustering, density-based clustering and grid-based clustering; also recent advances, such as subspace clustering, text clustering and data stream clustering. The second contribution by Emmanuel Udoh and Salim Bhuiyan “C-MICRA: A Tool for Clustering Microarray Data”, focuses on clustering as an important unsupervised method in the exploration of expression patterns in gene data arrays.


Doina Caragea and Vasant Honavar have contributed the article “Learning Classifiers from Distributed Data Sources” whose purpose is to precisely define the problem of learning classifiers from distributed data and summarize recent advances that have led to a solution to this problem. They describe a general strategy to transform standard machine learning algorithms—that assume centralized access to data in a single location—into algorithms to learn from distributed data.

The article “Differential Learning Expert System in Data Management” by Manjunath R. is devoted to problems related to knowledge acquisition for expert systems, and the analysis of plausible solutions for some of them. In this sense, the author exposes that a system using a rule-based expert system with an integrated connectionist network could benefit from the advantages of connectionist systems, regarding that machine-learning helps towards knowledge acquisition. The article presents a system based on rule-based expert system with neural networks which are able to perform a “learning from example” approach to extract rules from large data sets.

“Machine Learning as a Commonsense Reasoning Process” written by Xenia Naidenova concentrates on one of the most important tasks in database technology which is to combine the activities of inferring knowledge from data (data mining) and reasoning on acquired knowledge (query processing). The article includes a proposal of a unified model of commonsense reasoning, and also a demonstration showing that a large class of inductive machine learning (ML) algorithms can be transformed into the commonsense reasoning processes based on well-known deduction and induction logical rules.

“Machine Learning and Data Mining in Bioinformatics” is a contribution coauthored by George Tzanis, Christos Berberidis, and Ioannis Vlahavas. In this article, the authors review the exponential growth of biological data and the new questions these data have originated, due to recent technological advances. In particular, they focus on the mission of bioinformatics as a new and critical research
domain, which must provide the tools and use them to extract accurate and reliable information in order to gain new biological insights.

The contribution “Sequential Pattern Mining from Sequential Data” overviews sequential pattern discovery methods from discrete sequential data. Its author, Shigeaki Sakurai, focuses on sequential interestingness, which is an evaluation criterion of sequential patterns, highlighting that there are 7 types of time constraints that are the background knowledge related to the interests of analysts.

The field of scientometrics has been looking at the identification of co-authorship through network mapping. In a similar context, the paper entitled “From Chinese Philosophy to Knowledge Discovery in Databases: A Case Study: Scientometric Analysis” by Pei Liu explores the latent association of two authors, i.e. the collaboration between two researchers which has not yet occurred but might take place in the future. The author also shows how the concepts of Yuan (Interdependent arising), Kong (Emptiness), Shi (Energy) and Guanxi (Relationship) in Chinese philosophy contribute to understand ‘latent associations’, bringing in this way an original approach which could be applicable to the database research community.

Other contributions dealing with Data Mining, data warehousing and knowledge acquisition aspects can be found in the section Conceptual Modeling (Section I): “Schema Evolution Models and Languages for Multidimensional Data Warehouses” by Edgard Benitez-Guerrero, Ericka-Janet Rechy-Ramirez, “A Survey of Data Warehouse Model Evolution” by Cécile Favre, Fadila Bentayeb and Omar Boussaid, “Principles on Symbolic Data Analysis” by Héctor Oscar Nigro and Sandra Elizabeth González Císaro and three articles “Different Kinds of Hierarchies in Multidimensional Models”, “Spatial Data in Multidimensional Conceptual Models” and “Requirement Specification and Conceptual Modeling for Data Warehouses” by Elzbieta Malinowski, in the section Spatial and Temporal Databases (Section III): “Automatic Data Enrichment in GIS Through Condensate Textual Information” by Khaoula Mahmoudi, and Sami Faïz, “Similarity Search in Times Series” by Maria Kontaki, Apostolos N. Papadopoulos and Yannis Manolopoulos and “Current Approaches and Future Trends of Ontology-Driven Geographic Integration” by Agustina Buccella and Alejandra Cechich, in the section Ontologies (Section V): “Ontologies Application to Knowledge Discovery Process in Databases” by Héctor Oscar Nigro and Sandra Elizabeth González Císaro and in the section Physical Issues (Section VII): “Index and Materialized View Selection in Data Warehouses” by Kamel Aouiche and Jérôme Darmont, “Full-Text Manipulation in Databases” by László Kovács and Domonkos Tikk “Synopsis Data Structures for Representing, Querying, and Mining Data Streams” and “Innovative Access and Query Schemes for Mobile Databases and Data Warehouses” both by Alfredo Cuzzocrea.

Section VII: Physical Issues

The increasing number of database paradigms, database applications, types of data stored and database storing techniques leads to several new physical issues regarding storage requirements, information retrieval and query processing. New indexing techniques, document clustering, materialized views, commit protocols, data replications and crash recovery issues are partial but important answers to these concerns, among many others. This section contains several research reports and tutorials on the state of art of physical issues.

In “An Overview on Signature File Techniques”, Yangjun Chen presents an overview on recent relevant research results on information retrieval, mainly on the creation of database indexes which can be searched efficiently for the data under seeking. The focus of this article is on signature techniques.

The following contributions deal with efficiency on XML Databases. The article by Yangjun Chen, “On the Query Evaluation in XML Databases”, presents a new and efficient algorithm for XML query
processing, reducing the time and space needed to satisfy queries. In the article “XML Document Clustering”, Andrea Tagarelli provides a broad overview of the state-of-the-art and a guide to recent advances and emerging challenges in the research field of clustering XML documents. Besides basic similarities criteria based on structure of the document, the article focus is on the ability of clustering XML documents without assuming the availability of predefined XML schemas. Finally, “Indices in XML Databases”, a contribution by Hadj Mahboubi and Jérôme Darmont presents an overview of state-of-the-art XML indexes, discusses the main issues, tradeoffs and future trends in XML indexing and, since XML is gaining importance for representing business data for analytics, it also presents an index that the authors specifically developed for XML data warehouses.

In the article “XML Document Clustering”, Andrea Tagarelli provides a broad overview of the state-of-the-art and a guide to recent advances and emerging challenges in the research field of clustering XML documents. Besides basic similarities criteria based on structure of the document, the article focus is on the ability of clustering XML documents without assuming the availability of predefined XML schemas.

“Integrative Information Systems Architecture: Document & Content Management”, an article submitted by Len Asprey, Rolf Green, and Michael Middleton, overviews benefits of managing business documents and Web content within the context of an integrative information systems architecture which incorporates database management, document and Web content management, integrated scanning/imaging, workflow and capabilities of integration with other technologies.

The contribution by Kamel Aouiche and Jérôme Darmont, “Index and Materialized View Selection in Data Warehouses”, presents an overview of the major families of state-of-the-art index and materialized view selection methods; discusses the issues and future trends in data warehouse performance optimization, and focuses on data mining-based heuristics to reduce the selection problem complexity.

“Synopsis Data Structures for Representing, Querying, and Mining Data Streams”, contributed by Alfredo Cuzzocrea, provides an overview of state-of-the-art of synopsis data structures for data streams, making evident the benefits and limitations of each of them in efficiently supporting representation, query, and mining tasks over data streams.

“GR-OLAP: On Line Analytical Processing of GRid Monitoring Information”, the article contributed by Julien Gossa and Sandro Bimonte deals with the problem of management of Grid networks. The authors discuss recent advances in Grid monitoring, proposing the use of data warehousing and on line analytical processing to mine Grid monitoring information to get knowledge about the Grid networks characteristics.

“A Pagination Method for Indexes in Metric Databases”, a paper by Ana Villegas, Carina Ruano, and Norma Herrera, proposes an original strategy for metric databases whose index and/or data do not fit completely in the main memory. This strategy adapts the metric database regarding the capacity of the main memory, instead of adapting the index to be efficiently handled in secondary memory.

“SWIFT: A Distributed Real Time Commit Protocol”, an article submitted by Udai Shanker, Manoj Misra, and Anil K. Sarje, introduces a protocol to reduce the time to reach the commit in some specific situations, in the context of distributed databases.

“MECP: A Memory Efficient Real Time Commit Protocol”, coauthored by Udai Shanker, Manoj Misra, and Anil K. Sarje presents the problem of handling huge databases in the context of real time applications. In both situations, any saving in main memory usage becomes very important. In this article, the design of a distributed commit protocol which optimizes memory usage is presented.

The article “Self-Tuning Database Management Systems” has been contributed by Camilo Porto Nunes, Cláudio de Souza Baptista, and Marcus Costa Sampaio and addresses the issue of self-tuning DBMS, presenting a background on this topic followed by a discussion centered on performance, indexing and memory issues.
The article “Database Replication Approaches” contributed by Francesc Muñoz-Escoi, Hendrik Decker, José Armendáriz, and José González de Mendivil revise different approaches tackling the problem of database replication management. The authors analyze new replication techniques that were introduced for databases—as an evolution of the process replication approaches found in distributed systems.

“A Novel Crash Recovery Scheme for Distributed Real-Time Databases”, is a contribution by Yingyuan Xiao that reports research results into the crash recovery strategy area for distributed real-time main memory database systems (DRTMMDBS), including real-time logging scheme, local fuzzy checkpoint and dynamic recovery processing strategy.

The article “Querical Data Networks” (QDN) by Cyrus Shahabi and Farnoush Banaei-Kashani defines and characterizes QDNs as a new family of data networks with common characteristics and applications. It also reviews possible database-like architectures for QDNs as query processing systems and enumerates the most important QDN design principles. The authors also address the problem of effective data location for efficient query processing in QDNs, as the first step toward comprehending the vision of QDNs as complex distributed query-processing systems.

“On the Implementation of a Logic Language for NP Search and Optimization Problems” an article by Sergio Greco, Cristian Molinaro, Irina Trubitsyna, and Ester Zumpano, presents the logic language NP Datalog. It is a restricted version of DATALOG to formulate NP search and optimization problems which admits only controlled forms of negation such as stratified negation, exclusive disjunction and constraints and enables a simpler and intuitive formulation for search and optimization problems. In this contribution, a solution based on the rewriting of logic programs into constraint programming is proposed.

The article by Alfredo Cuzzocrea, “A Query-Strategy-Focused Taxonomy of P2P IR Techniques”, presents a taxonomy of the state-of-the-art of peer-to-peer (P2P) systems-information retrieval (IR) techniques, with emphasis on the query strategy used to retrieve information and knowledge from peers; and shows similarities and differences among the techniques.

In their contribution “Pervasive and Ubiquitous Computing Databases: Critical Issues and Challenges”, the authors Michael Zouboulakis and George Roussos offer a survey on the dual role that databases have to play in Pervasive and Ubiquitous Computing. In the short-term, they need to provide the mapping between physical and virtual entities and space in a highly distributed and heterogeneous environment while in the long term database management systems need to provide the infrastructure for the development of data-centric systems.

The following two contributions, written by the Christoph Bussler, deal with business integration. The former “Business-to-Business (B2B) Integration” surveys on how B2B integration is absolutely essential for business and organizations not only to stay competitive but also keep or even gain market share. The latter “Enterprise Application Integration (EAI)” by the same author surveys on current developments and critical issues of enterprise application integration (EAI) technologies, as they are essential for enterprises with more than one back end application system.

In a world in which globalization is increasingly integrating the economies and societies, products created in one nation are often marketed to a range of international consumers. Cross-border interactions on social and professional levels have been facilitated by the rapid diffusion of online media, however, different cultural expectations can cause miscommunication within this discourse paradigm. Localization has thus become an important aspect of today’s global economy. “The Role of Rhetoric in Localization and Offshoring”, a contribution by Kirk St.Amant, focuses on these issues, examining localization in offshoring practices that could affect database creation and maintenance.

In “Adaptive XML-to-Relational Storage Strategies”, Irena Mlynkova provides an overview of existing XML-to-relational storage strategies. This paper examines their historical development and provides a more detailed discussion of the currently most promising ones—the adaptive methods.
“Innovative Access and Query Schemes for Mobile Databases and Data Warehouses” authored by Alfredo Cuzzocrea presents a critical discussion on several aspects of mobile databases and data warehouses, along with a survey on state-of-the-art data-intensive mobile applications and systems. The Hand-OLAP system, a relevant instance of mobile OLAP systems is also described.

The paper “Full-Text Manipulation in Databases”, by László Kovács and Domonkos Tikk overviews issues and problems related to full-text search (FTS). The authors’ aim is to elucidate about the needs of users which usually require additional help to exploit the benefits of the functionalities of the FTS engines, such as: stemming, synonym and thesaurus based matching, fuzzy matching and Boolean operators. They also point out that current research focuses on solving the problem of covering new document formats, adapting the query to the user’s behavior, and providing an efficient FTS engine implementation.

“Bind but Dynamic Technique: The Ultimate Protection Against SQL Injections” a contribution by Ahmad Hammoud and Ramzi A. Haraty, explores on the risk and the level of damage that might be caused when web applications are vulnerable to SQL injections, and provides an efficient solution.

Other contributions dealing with Physical Issues can be found in the section Conceptual Modeling (Section I): “Document Versioning and XML in digital libraries” by M. Mercedes Martínez-González, in the section Spatial and Temporal Databases (Section III): “Spatio-Temporal Indexing Techniques” by Michael Vassilakopoulos and Antonio Corral and “Query processing in spatial databases” by Antonio Corral and Michael Vassilakopoulos, in the section Ontologies (Section V): “Mediation and Ontology-Based Framework for Interoperability” by Leonid Stoimenov and “Similarity Retrieval and Cluster Analysis Using R*-Trees” by Jiaxiong Pi, Yong Shi, and Zhengxin Chen and in the section Data Mining (Section VI): “Managing Temporal Data” by Abdullah Uz Tansel.

Summing up, this handbook offers an interesting set of articles about the state of the art of fundamental database concepts and a unique compilation of chapters about new technologies, current research trends, and challenging applications addressing the needs of present database and information systems.