Preface

Computer-based information technologies have been extensively used to help many organizations, private companies, and academic and education institutions manage their processes and information systems hereby become their nervous center. Information systems are used to manage data. The explosion of massive data sets created by businesses, science and governments necessitates intelligent and more powerful computing paradigms so that users can benefit from this data. This information needs to be summarized and synthesized to support effective problem solving and decision making. Various techniques have been proposed for intelligent data processing and analysis in the area of artificial intelligence (AI). Machine learning, for example, would implement various forms of learning, in particular mechanisms capable of inducing knowledge from examples or data.

Databases are designed to support the data storage, processing, and retrieval activities related to data management in information systems. Database management systems provide efficient task support and tremendous gain in productivity is hereby accomplished using these technologies. Database systems are the key to implementing data management. Data management requires database technique support. Database technology is typically application-oriented. With advances and in-deep applications of computer technologies, in particular, the extensive applications of Web technology in various areas, databases have become the repositories of large volumes of data. It is very critical to manage and use the worth data resource for effective problem solving and decision making. The research and development of intelligent databases are hereby emerging as a new discipline and are receiving increasing attention. The knowledge discovery in databases and data mining has witnessed it.

By means of database technology, large volumes of data can be stored in databases. Meanwhile large volumes of data in databases can be handled by means of AI technology and database systems should provide task support for such information processing. The next generation of information systems will be built based on intelligent databases to support various problem solving and decision making. So the study of intelligent databases is a field that must be investigated by academic

researchers together with developers and users both from database and AI areas. This book focuses on two major issues of intelligent databases, namely, intelligent information processing in databases and intelligent aspects of database systems, and presents the latest research and application results in intelligent databases. The different chapters in the book have been contributed by different authors and provide possible solutions for the different types of technological problems concerning intelligent databases.

Introduction

This book which consists of 11 chapters is organized into two major sections. The first section discusses the issues of intelligent information processing in databases in the first eight chapters. The next three chapters covering the intelligent aspects of database systems comprise the second section.

First of all, we take a look at the problems of the intelligent information processing in databases.

Most algorithms and approaches dealing with data mining in general and especially those focusing on the task of association rule mining have assumed all items to be only positively correlated, and looked only into the items that remained finally in a shopping basket. Very few works have proposed the existence of negative correlations between items, based though on the absence of items from transactions rather than on their actual removals. Kouris, Makris, and Tsakalidis look into mining that takes into consideration valuable information from rejected items and propose various alternatives for taking the specific items into account efficiently. The authors finally provide experimental evidence on the existence and significance of these items.

Outlier detection is an important research issue in data mining fields. The number of cells in the cell-based disk algorithm increases exponentially. The performance of this algorithm decreases dramatically with the increasing of the number of cells and data points. Zhao, Bao, Sun, and Yu find that there are many empty cells that are useless to outlier detection. They propose a novel index structure, called CD-Tree, in which only non-empty cells are stored, and a cluster technique is adopted to store the data objects in the same cell into linked disk pages. They test the performance of the proposed algorithms.

In the field of data mining, developing an efficient mining algorithm that can incrementally maintain discovered information as a database grows is quite important. Deletion of records in databases is commonly seen in real-world applications. Hong and Wang propose an incremental mining algorithm for maintenance of association rules as new transactions are inserted. The authors first review the maintenance of association rules from data insertion and then attempt to extend it to solve the data

deletion issue. The concept of pre-large itemsets is used to reduce the need for rescanning the original database and to save maintenance costs. A novel algorithm is proposed to maintain discovered association rules for deletion of records. The proposed algorithm doesn't need to rescan the original database until a number of records have been deleted.

Inductive databases have been proposed as general purpose databases to support the KDD process. Unfortunately, the heterogeneity of the discovered patterns and of the different conceptual tools used to extract them from source data makes difficult the integration in a unique framework. Meo and Psaila explore the feasibility of using XML as the unifying framework for inductive databases, and propose a new model, XML for data mining (XDM). They show the basic features of the model, based on the concepts of data item (source data and patterns) and statement (used to manage data and derive patterns). They make use of XML namespaces (to allow the effective coexistence and extensibility of data mining operators) and of XML-schema, by means of which they define the schema, the state and the integrity constraints of an inductive database.

Although data warehouse and geographical information system technologies are very useful in the decision making process, they are usually used separately. Integrating these two technologies coins new terms: spatial datawarehouse (SDW) and spatial OLAP (SOLAP). By using SOLAP, users may enhance their capacity to explore the underlying dataset once spatial methods incorporated into OLAP ones may be used. Sampaio et al. propose an integrated architecture for a SDW, including a formalized data model for SDW, a SQL extension query language which enables spatial roll-up and drill-down, optimization techniques which improve performance of complex spatial queries by pre-storing spatial aggregates, and a prototype, *MapWarehouse*, which validates the ideas proposed.

Data mining has the capability for classification, prediction, estimation, and pattern recognition by using manufacturing databases. Databases of manufacturing systems contain significant information for decision making, which could be properly revealed with the application of appropriate data mining techniques. Oke demonstrates the application of decision tree, a data mining tool, in the manufacturing system. Decision trees are employed for identifying valuable information in manufacturing databases. Practically, industrial managers would be able to make better use of manufacturing data at little or no extra investment in data manipulation cost. The author shows that it is valuable for managers to mine data for better and more effective decision making.

Zarri evokes the ubiquity and the importance of the so-called “narrative” information, showing that the usual ontological tools are unable to offer complete and reliable solutions for representing and exploiting this type of information. Then the author supplies some details about Narrative Knowledge Representation Language (NKRL), a fully-implemented knowledge representation and inferencing environment especially created for an “intelligent” exploitation of narrative knowledge.

The main innovation of NKRL consists in associating with the traditional ontologies of concepts an "ontology of events", in other words, a new sort of hierarchical organization where the nodes correspond to n-ary structures representing formally generic classes of elementary events like "move a physical object", "be present in a place" or "send/receive a message". More complex, second order tools based on the "reification" principle allow one to encode the "connectivity phenomena" like causality, goal, indirect speech, co-ordination, and subordination that, in narrative information, link together "elementary events". The chapter includes a description of the inference techniques proper to NKRL, and some information about the last developments of this language.

A major goal for database research has been the incorporation of additional semantics into the data model. Classical data models often suffer from their incapability of representing and manipulating imprecise and uncertain information that may occur in many real-world applications. Therefore, fuzzy set theory has been extensively applied to extend various data models and resulted in numerous contributions, mainly with respect to the popular relational model or to some related form of it. To satisfy the need of modeling complex objects with imprecision and uncertainty, recently many researches have been concentrated on fuzzy semantic (conceptual) and object-oriented data models. Ma reviews fuzzy database modeling technologies, including fuzzy conceptual data models and database models. Concerning fuzzy database models, fuzzy relational databases and fuzzy object-oriented databases are discussed, respectively.

In the second section, we see the intelligent aspects of database systems.

Wolff describes some of the kinds of "intelligence" that may be exhibited by an intelligent database system based on the SP theory of computing and cognition. The author introduces the SP theory and its main attractions as the basis for an intelligent database system: that it uses a simple but versatile format for diverse kinds of knowledge, that it integrates and simplifies a range of AI functions, and that it supports established database models when that is required. Then with examples and discussion, the author illustrates aspects of "intelligence" in the system: pattern recognition and information retrieval, several forms of probabilistic reasoning, the analysis and production of natural language, and the unsupervised learning of new knowledge.

Integrity constraints are a key tool for characterizing the well-formedness and semantics of the information contained in databases. In this regard, it is essential that intelligent database management systems provide their users with automatic support to effectively and efficiently maintain the semantic correctness of data with respect to the given integrity constraints. Martinenghi, Christiansen, and Decker give an overview of the field of efficient integrity checking and maintenance for relational as well as deductive databases. It covers both theoretical and practical aspects of integrity control, including integrity maintenance via active rules. The authors outline new lines of research, particularly with regard to two topics where a

strong impact for future developments can be expected: integrity in XML document collections and in distributed databases. Both pose a number of new and highly relevant research challenges to the database community.

Information systems, including their core databases need to meet changing user requirements and adhere to evolving business strategies. Traditional database evolution techniques focus on reacting to change to smoothly perform schema evolution operations and to propagate corresponding updates to the data as effectively as possible. Adopting such posterior solutions to such changes generates high costs in human resources and financial support. Bounif advocates an alternate solution: a predictive approach to database evolution. In this approach, ones anticipate future changes during the standard requirements analysis phase of schema development. The approach enables potential future requirements to be planned for, as well as the standard, determining what data is to be stored and what access is required. This preparation contributes significantly in the ability of the database schema to adapt to future changes and to estimate their relative costs.