Preface

Databases are designed to support the data storage, processing, and retrieval activities related to data management. The wide usage of databases in various applications has resulted in an enormous wealth of data, which populate various types of databases around the worlds. Ones can find many types of database systems, for example, relational databases, object-oriented databases, object-relational databases, deductive databases, parallel databases, distributed databases, multidatabase systems, Web databases, XML databases, multimedia databases, temporal/spatial databases, spatiotemporal databases, and uncertain databases. As a result, databases have become the repositories of large volumes of data.

Database query is closely related to data management. Database query processing is such a procedure that database management systems (DBMSs) obtain the information needed by the users from the databases according to users' requirements, and then provides them to the users after this useful information is organized. It is very critical to deal with the enormity and retrieve the worthwhile information for effective problem solving and decision making. It is especially true when a variety of database types, data types, and users' requirements, as well as large volumes of data, are available. The techniques of database queries are challenging today's database systems and promoting their evolvement. There is no doubt that database query systems play an important role in data management, and data management requires database query support.

The research and development of information queries over a variety of databases are receiving increasing attention. By means of query technology, large volumes of information in databases can be retrieved, and Information Systems are hereby built based on databases to support various problem solving and decision making. So database queries are the fields which must be investigated by academic researchers together with developers and users both from database and industry areas.

This book focuses on the following issues of advanced database query systems: the technologies and methodologies of database queries, XML and metadata queries, and applications of database query systems, aiming at providing a single account of technologies and practices in advanced database query systems. The objective of the book is to provide the state of the art information to academics, researchers and industry practitioners who are involved or interested in the study, use, design, and development of advanced and emerging database queries with ultimate aim to empower individuals and organizations in building competencies for exploiting the opportunities of the data and knowledge society. This book presents the latest research and application results in advanced database query systems. The different chapters in the book have been contributed by different authors and provide possible solutions for the different types of technological problems concerning database queries.

This book, which consists of fourteen chapters, is organized into three major sections. The first section discusses the technologies and methodologies of database queries, over the first eight chapters. The

next three chapters covering XML and metadata queries comprise the second section. The third section, containing the final three chapters, focuses on the design and applications of database query systems.

First of all, we take a look at the issues of the technologies and methodologies of database queries.

Web database queries are often exploratory. The users often find that their queries return too many answers and many of them may be irrelevant. Based on different kinds of user preferences, Xiangfu Meng, Li Yan and Z. M. Ma propose a novel categorization approach which consists of two steps. The first step analyzes query history of all users in the system offline and generates a set of clusters over the tuples, where each cluster represents one type of user preference. When a user issues a query, the second step presents to the user a category tree over the clusters generated in the first step such that the user can easily select the subset of query results matching his needs. The problem of constructing a category tree is a cost optimization problem and the authors develop heuristic algorithms to compute the min-cost categorization. The efficiency and effectiveness of their approach are demonstrated by experimental results.

Database systems are increasingly used for interactive and exploratory data retrieval. In such retrievals, user queries often result in too many answers, so users waste significant time and efforts sifting and sorting through these answers to find the relevant ones. Mounir Bechchi, Guillaume Raschia and Noureddine Mouaddib first review and discuss several research efforts that have attempted to provide users with effective and efficient ways to access databases. Then, they focus on a simple but useful strategy for retrieving relevant answers accurately and quickly without being distracted by irrelevant ones. They present a very recent but promising approach to quickly provide users with structured and approximate representations of users' query results, a must have for decision support systems. The underlying algorithm operates on pre-computed knowledge-based summaries of the queried data, instead of raw data themselves. Thus, this first-citizen data structure is also presented.

Alexandr Savinov describes a novel query language, called the concept-oriented query language (COQL), and demonstrates how it can be used for data modeling and analysis. The query language is based on a novel construct, called concept, and two relations between concepts, inclusion and partial order. Concepts generalize conventional classes and are used for describing domain-specific identities. This includes relation generalized inheritance and is used for describing hierarchical address spaces. Partial order among concepts is used to define two main operations: projection and de-projection. Savinov demonstrates how these constructs are used to solve typical tasks in data modeling and analysis such as logical navigation, multidimensional analysis, and inference.

Criteria that induce a Skyline naturally represent user's preference conditions useful to discard irrelevant data in large datasets. However, in the presence of high-dimensional Skyline spaces, the size of the Skyline can still be very large. To identify the best k points among the Skyline, the Top-k Skyline approach has been proposed. Marlene Goncalves and María-Esther Vidal describe existing solutions and propose to use the TKSI algorithm for the Top-k Skyline problem. TKSI reduces the search space by computing only a subset of the Skyline that is required to produce the top-k objects. In addition, the Skyline Frequency Metric is implemented to discriminate among the Skyline objects those that best meet the multidimensional criteria. They empirically study the quality of TKSI, and their experimental results show the TKSI may be able to speed up the computation of the Top-k Skyline in at least 50% percent with regards to the state-of-the-art solutions.

Janusz Kacprzyk, Guy De Tré, and Sławomir Zadrożny briefly present the concept of, a rationale for and various approaches to the use of fuzzy logic in flexible querying. They discuss first some historical developments, and then the main issues related to fuzzy querying. Next, they concentrate on fuzzy queries

with linguistic quantifiers, and discuss in more detail their FQUERY for Access fuzzy querying system. They indicate not only the straightforward power of that fuzzy querying system but its great potential as a tool to implement linguistic data summaries that may provide an ultimately human consistent way of data mining and data summarization. Also, they briefly mention the concept of bipolar queries that may reflect positive and negative preferences of the user, and may be a breakthrough in fuzzy querying. In the context of fuzzy querying and linguistic summarization they mention a considerable potential of their new recent proposals to explicitly use in linguistic data summarization some elements of natural language generation (NLG), and some natural language generation related elements of Halliday's systemic functional linguistics (SFL). They argue that this may be a promising direction for future research.

Gloria Bordogna *et al*. discuss the limitations of current temporal metadata in discovery services of Spatial Data Infrastructures (SDIs) and propose some solutions. They present their proposal of a formal and operational method to represent imperfect temporal metadata values and allow users to express flexible search conditions, i.e. tolerant to under-satisfaction. In doing so, discovery services can apply partial matching mechanisms between the "desired" metadata, expressed by the user, and the archived metadata: this would allow retrieving geodata in decreasing order of relevance to the user needs, as it usually occurs on the Web when using search engines. The proposal is finally illustrated with an example.

Ana Aguilera, José Tomás Cadenas and Leonid Tineo concentrate on incorporating the fuzzy capabilities to a relational database management system (RDBMS) of open source. The fuzzy capabilities include connectors, modifiers, comparators, quantifiers, and queries. The extensions consider a more flexible DDL and DML languages. The aim is to show the design and implementation details in the RDBMS PostgreSQL. For this, they design and implement a fuzzy query processor and fuzzy access mechanism. Also, they define and implement the physical fuzzy relational operators. They show the flow of a fuzzy query through the different modules (parser, planner, optimizer, and executor). They include some experimental results to demonstrate the performance of the proposal solution. These results show that the extensions do not decrease the performance of the RDBMS.

Awadhesh Kumar Sharma, A. Goswami, and D.K. Gupta investigate the problems in integration of fuzzy relational databases and extend the relational data model to support fuzzy multidatabases of type-2 that contain integrated fuzzy relational databases. The extended model is given the name fuzzy tuple source (FTS) relational data model which is provided with a set of FTS relational operations to manipulate the global relations called FTS relations from such fuzzy multidatabases. They propose and implement a full set of FTS relational algebraic operations capable of manipulating an extensive set of fuzzy relational multidatabases of type-2 that include fuzzy data values in their instances. To facilitate formulation of global fuzzy query over FTS relations in such fuzzy multidatabases, an appropriate extension to SQL can be done so as to get fuzzy tuple source structured query language (FTS-SQL).

The second section deals with the issues of XML and metadata queries.

Tadeusz Pankowski addresses the problem of data integration in a P2P environment, where each peer stores schema of its local data, mappings between the schemas, and some schema constraints. The goal of the integration is to answer queries formulated against a chosen peer. The answer must consist of data stored in the queried peer as well as data of its direct and indirect partners. Pankowski focuses on defining and using mappings, schema constraints, query propagation across the P2P system, and query answering in such scenario. Schemas, mappings, constraints (functional dependencies) and queries are all expressed using a unified approach based on tree-pattern formulas. He discusses how functional dependencies can be exploited to increase information content of answers (by discovering missing values)

and to control merging operations and propagation strategies. He proposes algorithms for translating high-level specifications of mappings and queries into XQuery programs, and shows how the discussed method has been implemented in SixP2P (or 6P2P) system.

Significant research efforts in the Semantic Web community have recently been directed toward the representation and reasoning with fuzzy ontologies. Description logics (DLs) are the logical foundations of standard Web ontology languages. Conjunctive queries are deemed as an expressive reasoning service for DLs. Jingwei Cheng, Z. M. Ma, and Li Yan focus on fuzzy (threshold) conjunctive queries over knowledge bases encoding in fuzzy DL SHIF(**D**), the logic counterpart of fuzzy OWL Lite language. They show decidability of fuzzy query entailment in this setting by providing a corresponding tableau-based algorithm. Also they show data complexity for answering fuzzy conjunctive queries in fuzzy SHIF(**D**) is in coNP, as long as only simple roles occur in the query. Regarding combined complexity, they prove a co3NExpTime upper bound in the size of the knowledge base and the query.

The Resource Description Framework (RDF) is a flexible model for representing information about resources in the Web. With the increasing amount of RDF data which is becoming available, efficient and scalable management of RDF data has become a fundamental challenge to achieve the Semantic Web vision. The RDF model has attracted attentions in the database community, and many researchers have proposed different solutions to store and query RDF data efficiently. Sherif Sakr and Ghazi Al-Naymat concentrate on using relational query processors to store and query RDF data. They give an overview of the different approaches and classify these approaches according to the storage and query evaluation strategies.

In the third section, we see the design and application aspects of database query systems.

Relational Algebra (RA) and structured query language (SQL) are supposed to have a bijective relationship by having the same expressive power. That is, each operation in SQL can be mapped to one RA equivalent and vice versa. RA has an explicit relational division symbol (÷) whereas SQL does not have a corresponding explicit division keyword. Division is implemented using a combination of four core operations, namely cross product, difference, selection, and projection. The work described by Eric Draken, Shang Gao, and Reda Alhajj is intended to provide SQL expression equivalent to explicit relational algebra division (with static divisor). The goal is to implement a SQL query rewriter in Java which takes as input a divide grammar and rewrites it to an efficient query using current SQL keywords. The developed approach could be adapted as front-end or as a wrapper to existing SQL query system.

Recently, there has been a lot of interest in the application of graphs in different domains. Graphs have been widely used for data modeling in different application domains such as: chemical compounds, protein networks, social networks, and Semantic Web. Given a query graph, the task of retrieving related graphs as a result of the query from a large graph database is a key issue in any graph-based application. This has raised a crucial need for efficient graph indexing and querying techniques. Sherif Sakr and Ghazi Al-Naymat provide an overview of different techniques for indexing and querying graph databases. They also give an overview of several proposals of graph query language. Finally, they provide a set of guidelines for future research directions.

Multimedia objects–such as images, audio, and video–do not present the total ordering relationship, so the relational operators are not suitable to compare them. Therefore, similarity queries are the most useful, and often the only types of queries adequate to search multimedia objects stored in a database. Unfortunately, the ubiquitous query language SQL–the most widely employed language in Database Management Systems (DBMS)–does not provide effective support for similarity queries. Maria Camila

Nardini Barioni *et al*. present an already validated strategy that adds similarity queries to SQL, supporting a powerful set of similarity operators. They also describe techniques to store and retrieve multimedia objects in an efficient way and show existing DBMS alternatives to executing similarity queries over multimedia data.

Li Yan
Northeastern University, China

Zongmin Ma
Northeastern University, China