Preface

Spatial data management has been an active area of intensive research for more than two decades. In order to support spatial objects in a database system several issues should be taken into consideration including spatial data models, indexing mechanisms, efficient query processing, and cost models. One of the most influential access methods in the area is the R-tree structure proposed by Guttman in 1984 as an effective and efficient solution to index rectangular objects in VLSI design applications. Since then several variations of the original structure have been proposed to provide more efficient access, handle objects in high-dimensional spaces, support concurrent accesses, support I/O and CPU parallelism, and support efficient bulk loading. It seems that due to the modern demanding applications and after academia paved the way, recently the industry has recognized the use and necessity of R-trees. The simplicity of the structure and its resemblance to the B-tree allowed developers to easily incorporate the structure into existing database management systems to support spatial query processing. In this book we provide an extensive survey of the R-tree evolution, studying the applicability of the structure and its variations to efficient query processing, accurate proposed cost models, and implementation issues like concurrency control and parallelism. Based on the observation that "space is everywhere", we anticipate that we are in the beginning of the era of the "ubiquitous R-tree" in an analogous manner as B-trees were considered 25 years ago.

Book Organization

The book contains nine chapters organized in four parts. We tried to keep each chapter self-contained to provide maximum reading flexibility.

The first part of the book contains some fundamental issues and is composed of three chapters. Chapter 1 is the introductory chapter. In this chapter we give a brief introduction to the area, present the basic notations and the corresponding descriptions, and present the original R-tree access method, which is the root of the family tree of access methods presented in the book. Chapter 2 is devoted to the description of the most promising dynamic variations of the R-tree. These methods support insertions, deletions, and updates and therefore can be effectively used in a dynamic environment. Static variations

of the R-tree are given in Chapter 3. These variations are optimized taking into consideration that the dataset to be organized is given a priori.

The second part of the book is composed of two chapters and covers query processing techniques that have been proposed to operate with R-trees. Chapter 4 studies fundamental query types such as range queries, nearest-neighbor queries, and spatial join queries. Each method is studied in detail, and the corresponding algorithm is given in pseudo-code where appropriate. Chapter 5 explores more complex query types such as categorical range queries, multi-way spatial joins, closest-pair queries, incremental processing, and approximate query techniques. These queries are characterized by higher computation costs and greater complexity than the fundamental ones, and therefore are covered separately.

The adaptation of the R-tree to modern application domains is discussed in the third part, which comprises two chapters. Chapter 6 studies the application of R-tree-like access methods to spatiotemporal database systems. The fundamental characteristic of these systems is that they handle temporal information in addition to the spatial properties of objects. Chapter 7 discusses the use of R-trees in multimedia databases, data warehouses, and data mining tasks. The exploitation of the R-tree by the aforementioned domains has proven very promising to faster algorithms and query processing techniques, taking into consideration the complexity of objects and the computationally intensive operations required.

The last part of the book comprises two chapters. Chapter 8 studies query optimization issues for R-tree based query processing. Formulae are given for various query types that estimate the corresponding cost of the operation. These formulae are valuable for cost-based query optimization and selectivity estimation in modern database systems. Chapter 9 discusses some implementation issues regarding the R-tree access method. Many DBMS vendors have incorporated the R-tree as an indexing technique for non-traditional objects. Moreover, several research prototypes have implemented the R-tree to index spatial or multi-dimensional objects.

The Epilogue at the end of the book summarizes our work and gives some directions for future research in the area.

Intended Audience

We believe that this book (or parts of it) will be valuable to course instructors, undergraduate and postgraduate students studying access methods for advanced applications. Moreover, it will be a valuable companion to researchers and professionals working with access methods, because it presents in detail a broad range of concepts and techniques related to indexing, query processing, query optimization, and implementation. Finally, practitioners working in the development of access methods and database systems can use this book as a reference.

How to Study This Book

The order of presentation is the proposed reading order of the material. However, according to the reader's expertise in the area, it is possible to focus directly on the topic of interest by studying the corresponding chapters. Expert readers could skip the first part of the book, whereas non-experts are encouraged to start from the beginning. Evidently, if the reader wishes to study more details with respect to a specific topic, then the corresponding references should be consulted.

More precisely, undergraduate students can focus on Chapters 1 through 4 and Chapter 9 to grasp the main characteristics of the R-tree and related access methods, and to understand how query processing is performed in a spatial database system. Postgraduate students will find Chapters 5, 6, and 8 motivating for further research in the area, because efficientquery processing and optimization techniques are active research fields. Course instructors and researchers should study all the material to select parts of the book required for class presentation or further research in the area.

Acknowledgments

A significant number of the papers contributed by the authors of this effort were the outcome of research funded by the EU Chorochronos program.

We would like to thank the co-authors of our papers: P. Bozanis, S. Brakatsoulas, A. Corral, M. Egenhofer, C. Faloutsos, C. Gurret, Y. Karydis, N. Mamoulis, E. Nardelli, M. Nascimento, D. Papadias, D. Pfoser, G. Proietti, M. Ranganathan, P. Rigaux, J.R.O. Silva, M. Scholl, T. Sellis, E. Stefanakis, V. Vasaitis, M. Vassilakopoulos, and M. Vazirgiannis.

There are also several people who helped significantly in the preparation of this book. We take the opportunity to thank Antonio Corral for his contribution in approximate queries cost models for distance joins, and Elias Frentzos for his contribution in query optimization issues.

Finally, we would like to thank Catherine Drury and Michael Koy from Springer, for their great help and support toward the completion of this project. Their comments and suggestions were very helpful in improving the readability, organization, and overall view of the book.

We hope that this book will be a valuable reference to the expert reader and a motivating companion for the non-expert who whishes to study the theory and applications of the R-tree access method and its variations.

Yannis Manolopoulos, Thessaloniki, Greece
Alexandros Nanopoulos, Thessaloniki, Greece
Apostolos N. Papadopoulos, Thessaloniki, Greece
Yannis Theodoridis, Athens, Greece

http://www.springer.com/978-1-85233-977-7

R-Trees: Theory and Applications

Manolopoulos, Y.; Nanopoulos, A.; Papadopoulos, A.N.;
Theodoridis, Y.

2006, XIX, 194 p. 77 illus., Hardcover
ISBN: 978-1-85233-977-7