IDEA GROUP PUBLISHING
701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com
ITB10784

Chapter VII

Spatial Joins: Algorithms, Cost Models and Optimization

Nikos Mamoulis, University of Hong Kong, Hong Kong

Yannis Theodoridis, University of Piraeus, Greece

Dimitris Papadias,
Hong Kong University of Science and Technology, Hong Kong

Abstract

This chapter describes algorithms, cost models and optimization techniques for spatial joins. Joins are among the most common queries in Spatial Database Management Systems. Due to their importance and high processing cost, a number of algorithms have been proposed covering all possible cases of indexed and non-indexed inputs. We first describe some popular methods for processing binary spatial joins and provide models for selectivity and cost estimation. Then, we discuss evaluation of multiway

spatial joins by integrating binary algorithms and synchronous tree traversal. Going one step further, we show how analytical models can be used to combine the various join operators in optimal evaluation plans. The chapter can serve as a comprehensive reference text to the researcher who wants to learn about this important spatial query operator and to the developer who wants to include spatial query processing modules in a Database System.

Introduction

Spatial database systems (Güting, 1994) manage large collections of multidimensional data which, apart from conventional features, include special characteristics such as position and extent. That there is no total ordering of objects in space that preserves proximity renders conventional indexes, such as B^+-trees, inapplicable to spatial databases. As a result, a number of *spatial access methods* have been proposed (Gaede & Günther, 1998). A very popular method, used in several commercial systems (for example, Informix and Oracle), is the R-tree (Guttman, 1994), which can be thought of as an extension of B^+-tree in multi-dimensional space. R-trees index object approximations, usually minimum bounding rectangles (MBRs), providing a fast *filter step* that excludes all objects that cannot satisfy a query. A subsequent *refinement step* uses the geometry of the candidate objects (that is, the output of the filter step) to dismiss false hits and retrieve the actual solutions. The R-tree and its variations have been applied to efficiently answer several query types, including spatial selections, nearest neighbors and spatial joins.

As in relational databases, joins play an important role in effective spatial query processing. A *binary* (that is, *pairwise*) spatial join combines two datasets with respect to a spatial predicate (usually *overlap*/*intersect*). A typical example is "find all pairs of cities and rivers that intersect." For instance, in Figure 1 the result of the join between the set of cities $\{c_1, c_2, c_3, c_4, c_5\}$ and rivers $\{r_1, r_2\}$, is $\{(r_1, c_1), (r_2, c_2), (r_2, c_5)\}$.

The query in this example is a spatial *intersection* join. In the general case, the join predicate could be a combination of *topological*, *directional* and *distance* spatial relations. Apart from the intersection join, variants of the *distance* join have received considerable attention because they find application in data analysis tasks (for example, data mining and clustering). Given two sets R and S of spatial objects (or multidimensional points) and a distance function *dist*(), the *ε-distance* join (or else *similarity* join) (Koudas & Sevcik, 2000) returns the

Related Content

Enterprise Application Integration (EAI)
Christoph Bussler (2009). *Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends (pp. 837-843).*
www.irma-international.org/chapter/enterprise-application-integration-eai/20769

Applying UML for Designing Multidimensional Databases and OLAP Applications
Juan Trujillo, Sergio Lujan-Moraand Il-Yeol Song (2003). *Advanced Topics in Database Research, Volume 2 (pp. 13-36).*
www.irma-international.org/chapter/applying-uml-designing-multidimensional-databases/4339

Incorporating I Ching Knowledge Into Prediction Task via Data Mining
Wenjie Liu, Sai Chen, Guoyao Huang, Lingfeng Lu, Huakang Liand Guozi Sun (2023). *Journal of Database Management (pp. 1-16).*
www.irma-international.org/article/incorporating-i-ching-knowledge-into-prediction-task-via-data-mining/322097

INTECoM: An Integrated Approach to the Specification and Design of Information Requirements
Clare Atkins (2001). *Developing Quality Complex Database Systems: Practices, Techniques and Technologies (pp. 240-260).*
www.irma-international.org/chapter/intecom-integrated-approach-specification-design/8278

Maintaining Mappings between Conceptual Models and Relational Schemas
Yuan An, Xiaohua Huand Il-Yeol Song (2010). *Journal of Database Management (pp. 36-68).*
www.irma-international.org/article/maintaining-mappings-between-conceptual-models/43729