Scalable Processing of Spatial-Keyword Queries

Synthesis Lectures on Data Management

Synthesis Lectures on Data Management is edited by H.V. Jagadish of the University of Michigan. The series publishes 80–150 page publications on topics pertaining to data management. Topics include query languages, database system architectures, transaction management, data warehousing, XML and databases, data stream systems, wide scale data distribution, multimedia data management, data mining, and related subjects.

Deep Web Query Interface Understanding and Integration
Eduard C. Dragut, Weiyi Meng, and Clement T. Yu
2012

P2P Techniques for Decentralized Applications
Esther Pacitti, Reza Akbarinia, and Manal El-Dick
2012

Query Answer Authentication
HweeHwa Pang and Kian-Lee Tan
2012

Declarative Networking
Boon Thau Loo and Wenchao Zhou
2012

Full-Text (Substring) Indexes in External Memory
Marina Barsky, Ulrike Stege, and Alex Thomo
2011

Spatial Data Management
Nikos Mamoulis
2011

Database Repairing and Consistent Query Answering
Leopoldo Bertossi
2011

Managing Event Information: Modeling, Retrieval, and Applications
Amarnath Gupta and Ramesh Jain
2011

Fundamentals of Physical Design and Query Compilation
David Toman and Grant Weddell
2011

Methods for Mining and Summarizing Text Conversations
Giuseppe Carenini, Gabriel Murray, and Raymond Ng
2011

Probabilistic Databases
Dan Suciu, Dan Olteanu, Christopher Ré, and Christoph Koch
2011

Peer-to-Peer Data Management
Karl Aberer
2011

Probabilistic Ranking Techniques in Relational Databases
Ihab F. Ilyas and Mohamed A. Soliman
2011

Uncertain Schema Matching
Avigdor Gal
2011

Fundamentals of Object Databases: Object-Oriented and Object-Relational Design
Suzanne W. Dietrich and Susan D. Urban
2010

Advanced Metasearch Engine Technology
Weiyi Meng and Clement T. Yu
2010

Web Page Recommendation Models: Theory and Algorithms
Sule Gündüz-Ögüdücü
2010

Multidimensional Databases and Data Warehousing
Christian S. Jensen, Torben Bach Pedersen, and Christian Thomsen
2010

Database Replication
Bettina Kemme, Ricardo Jimenez-Peris, and Marta Patino-Martinez
2010

Relational and XML Data Exchange
Marcelo Arenas, Pablo Barcelo, Leonid Libkin, and Filip Murlak
2010

User-Centered Data Management
Tiziana Catarci, Alan Dix, Stephen Kimani, and Giuseppe Santucci
2010

Data Stream Management
Lukasz Golab and M. Tamer Özsu
2010

Access Control in Data Management Systems
Elena Ferrari
2010

An Introduction to Duplicate Detection
Felix Naumann and Melanie Herschel
2010

Scalable Processing of Spatial-Keyword Queries

Ahmed R. Mahmood and Walid G. Aref

ISBN: 978-1-031-00719-2 paperback
ISBN: 978-3-O31-01867-1 ebook
ISBN: 978-1-031-00094-2 hardcover

DOI 10.1007/978-3-O31-01867-1

A Publication in the Springer scries
SYNTHESIS LECTURES ON DATA MANAGEMENT

Lecture #56
Series Editor: H.V. Jagadish, *University of Michigan*
Founding Editor: M. Tamer Özsu, *University of Waterloo*
Series ISSN
Print 2153-5418 Electronic 2153-5426

Scalable Processing of Spatial-Keyword Queries

Ahmed R. Mahmood and Walid G. Aref
Purdue University

SYNTHESIS LECTURES ON DATA MANAGEMENT #56

ABSTRACT

Text data that is associated with location data has become ubiquitous. A tweet is an example of this type of data, where the text in a tweet is associated with the location where the tweet has been issued. We use the term spatial-keyword data to refer to this type of data. Spatial-keyword data is being generated at massive scale. Almost all online transactions have an associated spatial trace. The spatial trace is derived from GPS coordinates, IP addresses, or cell-phone-tower locations. Hundreds of millions or even billions of spatial-keyword objects are being generated daily. Spatial-keyword data has numerous applications that require efficient processing and management of massive amounts of spatial-keyword data.

This book starts by overviewing some important applications of spatial-keyword data, and demonstrates the scale at which spatial-keyword data is being generated. Then, it formalizes and classifies the various types of queries that execute over spatial-keyword data. Next, it discusses important and desirable properties of spatial-keyword query languages that are needed to express queries over spatial-keyword data. As will be illustrated, existing spatial-keyword query languages vary in the types of spatial-keyword queries that they can support.

There are many systems that process spatial-keyword queries. Systems differ from each other in various aspects, e.g., whether the system is batch-oriented or stream-based, and whether the system is centralized or distributed. Moreover, spatial-keyword systems vary in the types of queries that they support. Finally, systems vary in the types of indexing techniques that they adopt. This book provides an overview of the main spatial-keyword data-management systems (SKDMSs), and classifies them according to their features. Moreover, the book describes the main approaches adopted when indexing spatial-keyword data in the centralized and distributed settings. Several case studies of SKDMSs are presented along with the applications and query types that these SKDMSs are targeted for and the indexing techniques they utilize for processing their queries.

Optimizing the performance and the query processing of SKDMSs still has many research challenges and open problems. The book concludes with a discussion about several important and open research-problems in the domain of scalable spatial-keyword processing.

KEYWORDS

spatial-keyword, indexing, systems, big data, query processing

To my wife, my parents, and my children.

Ahmed R. Mahmood

To my beloved parents, professors Safaa Elhifni and Galal Aref, and my lovely grandchildren, Hana and Walid.

Walid G. Aref

Contents

Preface

Scalable processing of spatial-keyword data is an important problem that is being studied actively by database researchers and industrial practitioners. This book targets developers and researchers that are interested in the design of scalable spatial-keyword data management systems (SKDMSs). The book surveys and classifies the main spatial-keyword queries, query languages, indexing techniques, and processing models, and aims to aid developers and researchers improve the scalability, efficiency, and performance of SKDMSs. There are many open research problems in the area of spatial-keyword processing. In this book, we highlight some of these problems to motivate the engagement of database designers and researchers in addressing them. It does not require prior knowledge about spatial-keyword processing or systems. Basic knowledge of database systems and data indexing techniques are all that are needed.

Ahmed R. Mahmood and Walid G. Aref
January 2019

Acknowledgments

Walid G. Aref's research has been partially supported by the National Science Foundation under Grant III-1815796.