

# **Query Processing over Uncertain Databases**

# Synthesis Lectures on Data Management

Editor

**M. Tamer Özsu**, *University of Waterloo*

Synthesis Lectures on Data Management is edited by Tamer Özsu of the University of Waterloo. The series will publish 50- to 125 page publications on topics pertaining to data management. The scope will largely follow the purview of premier information and computer science conferences, such as ACM SIGMOD, VLDB, ICDE, PODS, ICDT, and ACM KDD. Potential topics include, but not are limited to: 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.

Query Processing over Uncertain Databases

Lei Chen and Xiang Lian

2012

Semantics Empowered Web 3.0: Managing Enterprise, Social, Sensor, and Cloud-based Data and Services for Advanced Applications

Amit Sheth and Krishnaprasad Thirunarayan

2012

Data Management in the Cloud: Challenges and Opportunities

Divyakant Agrawal, Sudipto Das, and Amr El Abbadi

2012

Foundations of Data Quality Management

Wenfei Fan and Floris Geerts

2012

Incomplete Data and Data Dependencies in Relational Databases

Sergio Greco, Cristian Molinaro, and Francesca Spezzano

2012

Business Processes: A Database Perspective

Daniel Deutch and Tova Milo

2012

Data Protection from Insider Threats

Elisa Bertino

2012

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

Privacy-Preserving Data Publishing: An Overview

Raymond Chi-Wing Wong and Ada Wai-Chee Fu

2010

Keyword Search in Databases

Jeffrey Xu Yu, Lu Qin, and Lijun Chang

2009

© Springer Nature Switzerland AG 2022

Reprint of original edition © Morgan & Claypool 2013

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopy, recording, or any other except for brief quotations in printed reviews, without the prior permission of the publisher.

Query Processing over Uncertain Databases

Lei Chen and Xiang Lian

ISBN: 978-3-031-00768-2      paperback

ISBN: 978-3-031-01896-1      ebook

DOI 10.1007/978-3-031-01896-1

A Publication in the Springer series

*SYNTHESIS LECTURES ON DATA MANAGEMENT*

Lecture #33

Series Editor: M. Tamer Özsu, *University of Waterloo*

Series ISSN

Synthesis Lectures on Data Management

Print 2153-5418    Electronic 2153-5426

# Query Processing over Uncertain Databases

Lei Chen

Hong Kong University of Science and Technology

Xiang Lian

University of Texas – Pan American

*SYNTHESIS LECTURES ON DATA MANAGEMENT #33*

## **ABSTRACT**

Due to measurement errors, transmission lost, or injected noise for privacy protection, uncertainty exists in the data of many real applications. However, query processing techniques for deterministic data cannot be directly applied to uncertain data because they do not have mechanisms to handle the data uncertainty. Therefore, efficient and effective manipulation of uncertain data is a practical yet challenging research topic. In this book, we start from the data models for imprecise and uncertain data, move on to defining different semantics for queries on uncertain data, and finally discuss the advanced query processing techniques for various probabilistic queries in uncertain databases. The book serves as a comprehensive guideline for query processing over uncertain databases.

## **KEYWORDS**

uncertain databases, possible worlds, attribute correlation, tuple correlation, spatial pruning, probabilistic pruning, refinement

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Uncertain Data Management	1
1.2	Real Applications of Uncertain Data Management	2
1.3	Classifications of Uncertain Data	5
1.4	Major Challenges of Uncertain Data Management	7
1.5	Outline	8
<b>2</b>	<b>Uncertain Data Models</b>	<b>9</b>
2.1	Uncertain Databases	9
2.2	Possible Worlds Semantics	10
2.3	Independent Uncertainty Model	11
2.4	Correlated Uncertainty Model	13
2.5	Uncertainty Model With Local Correlations	16
2.6	Exercises	19
<b>3</b>	<b>Spatial Query Semantics over Uncertain Data Models</b>	<b>21</b>
3.1	A Warm-Up Example of Probabilistic Range Query	21
3.2	Probabilistic Nearest Neighbor Queries	23
3.2.1	Motivation	23
3.2.2	Background	25
3.2.3	Definition of PNN	26
3.3	Probabilistic Reverse Nearest Neighbor Queries	27
3.3.1	Motivation	27
3.3.2	Background	29
3.3.3	Definition of PRNN	30
3.4	Probabilistic Reverse Skyline Queries	31
3.4.1	Motivation	31
3.4.2	Definition of PRS	35
3.5	Exercises	37

<b>4</b>	<b>Spatial Query Processing over Uncertain Databases</b> .....	<b>39</b>
4.1	General Framework .....	39
4.1.1	The Filter-and-Refine Framework .....	39
4.1.2	Highlights of Basic Pruning Techniques .....	41
4.2	Spatial Pruning .....	41
4.2.1	Spatial Pruning for Probabilistic Nearest Neighbor Queries .....	41
4.2.2	Spatial Pruning for Probabilistic Reverse Nearest Neighbor Queries ....	42
4.2.3	Spatial Pruning for Probabilistic Reverse Skyline Queries .....	50
4.3	Probabilistic Pruning .....	53
4.3.1	Probabilistic Pruning for Probabilistic Nearest Neighbor Queries .....	55
4.3.2	Probabilistic Pruning for Probabilistic Reverse Nearest Neighbor Queries .....	56
4.3.3	Probabilistic Pruning for Probabilistic Reverse Skyline Queries .....	58
4.4	Refinement .....	62
4.4.1	Refinement for Probabilistic Nearest Neighbor Queries .....	62
4.4.2	Refinement for Probabilistic Reverse Nearest Neighbor Queries .....	64
4.4.3	Refinement for Probabilistic Reverse Skyline Queries .....	65
4.5	Query Procedure .....	67
4.5.1	Probabilistic Nearest Neighbor Query Answering .....	68
4.5.2	Probabilistic Reverse Nearest Neighbor Query Answering .....	68
4.5.3	Probabilistic Reverse Skyline Query Answering .....	72
<b>5</b>	<b>Conclusion</b> .....	<b>79</b>
<b>A</b>	<b>Steps to Compute the Minimum Distance</b> .....	<b>81</b>
	<b>Bibliography</b> .....	<b>83</b>
	<b>Authors' Biographies</b> .....	<b>91</b>