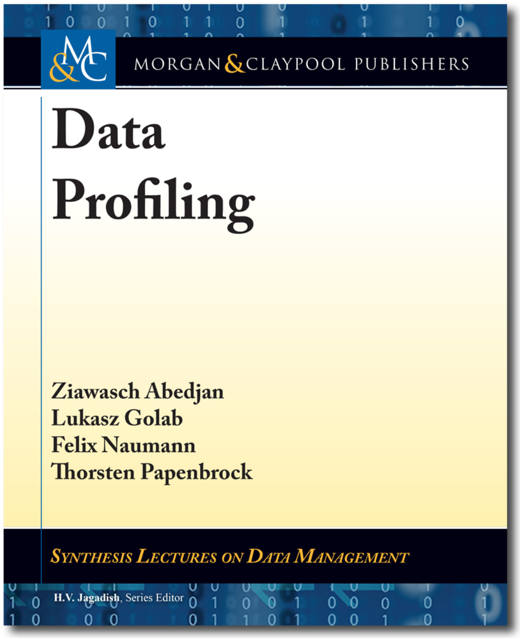
Classifying various types of profilable metadata, discussing popular data profiling tasks, and surveying state-of-the-art profiling algorithms.


MORGAN & CLAYPOOL PUBLISHERS

Data Profiling

Ziawasch Abedjan
Lukasz Golab
Felix Naumann
Thorsten Papenbrock

SYNTHESIS LECTURES ON DATA MANAGEMENT

H.V. Jagadish, Series Editor

Data Profiling

Ziawasch Abedjan, *Technische Universitat Berlin*
Lukasz Golab, *University of Waterloo*
Felix Naumann, *Hasso Plattner Institute, University of Potsdam*
Thorsten Papenbrock, *Hasso Plattner Institute, University of Potsdam*

Paperback ISBN: 9781681734460 • eBook ISBN: 9781681734477
Hardcover ISBN: 9781681734484 • November, 2018 • 154 pages
Paperback: $64.95 • eBook: $51.96 • Combo: $81.19
Hardcover: $84.95 • Hardcover Combo: $106.19

Data profiling refers to the activity of collecting data about data, i.e., metadata. Most IT professionals and researchers who work with data have engaged in data profiling, at least informally, to understand and explore an unfamiliar dataset or to determine whether a new dataset is appropriate for a particular task at hand. Data profiling results are also important in a variety of other situations, including query optimization, data integration, and data cleaning. Simple metadata are statistics, such as the number of rows and columns, schema and datatype information, the number of distinct values, statistical value distributions, and the number of null or empty values in each column. More complex types of metadata are statements about multiple columns and their correlation, such as candidate keys, functional dependencies, and other types of dependencies.

This book provides a classification of the various types of profilable metadata, discusses popular data profiling tasks, and surveys state-of-the-art profiling algorithms. While most of the book focuses on tasks and algorithms for relational data profiling, we also briefly discuss systems and techniques for profiling non-relational data such as graphs and text. We conclude with a discussion of data profiling challenges and directions for future work in this area.

CONTENTS

MORGAN & CLAYPOOL
PUBLISHERS
www.morganclaypoolpublishers.com
info@morganclaypool.com
Find Print, eBooks, and check for
Institutional Access all in one place.

Print & eBooks at http://store.morganclaypool.com