

Data Matching

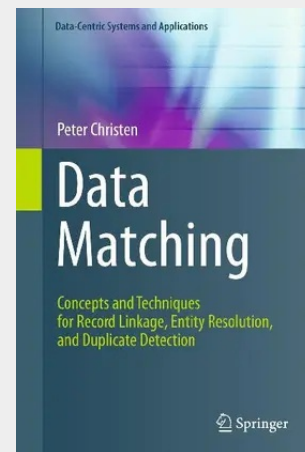
Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection

Data matching (also known as record or data linkage, entity resolution, object identification, or field matching) is the task of identifying, matching and merging records that correspond to the same entities from several databases or even within one database. Based on research in various domains including applied statistics, health informatics, data mining, machine learning, artificial intelligence, database management, and digital libraries, significant advances have been achieved over the last decade in all aspects of the data matching process, especially on how to improve the accuracy of data matching, and its scalability to large databases.

Peter Christen's book is divided into three parts: Part I, "Overview", introduces the subject by presenting several sample applications and their special challenges, as well as a general overview of a generic data matching process. Part II, "Steps of the Data Matching Process", then details its main steps like pre-processing, indexing, field and record comparison, classification, and quality evaluation. Lastly, part III, "Further Topics", deals with specific aspects like privacy, real-time matching, or matching unstructured data. Finally, it briefly describes the main features of many research and open source systems available today.

By providing the reader with a broad range of data matching concepts and techniques and touching on all aspects of the data matching process, this book helps researchers as well as students specializing in data quality or data matching aspects to familiarize themselves with recent research advances and to identify open research challenges in the area of data matching. To this end, each chapter of the book includes a final section that provides pointers to further background and research material. Practitioners will better understand the current state of the art in data matching as well as the internal workings and limitations of current systems. Especially, they will learn that it is often not feasible to simply implement an existing off-the-shelf data matching system without substantial adaption and customization. Such practical considerations are discussed for each of the major steps in the data matching process.

Data matching (also known as record or data linkage, entity resolution, object identification, or field matching) is the task of identifying, matching and merging records that correspond to the same entities from several databases or even within one database. Based on research in various domains including applied statistics, health informatics, data mining, machine learning, artificial intelligence, database management, and digital libraries, significant advances have been achieved over the last decade in all aspects of the data matching process, especially on how to improve the accuracy of data matching, and its scalability to large databases. Peter Christen's book is divided into three parts: Part I, "Overview", introduces the subject by presenting several sample applications and their special challenges, as well as a general overview of a generic data matching process. Part II, "Steps of the Data Matching Process", then details its main steps like pre-processing, indexing, field and record comparison, classification, and quality evaluation. Lastly, part III, "Further Topics", deals with specific aspects like privacy, real-time matching, or matching unstructured data. Finally, it briefly describes the main features of many research and open source systems available today. By providing the reader with a broad range of data matching concepts and techniques and touching on all aspects of the data matching process, this book helps researchers as well as students specializing in data quality or data matching aspects to familiarize themselves with recent research advances and to identify open research challenges in the area of data matching. To this end, each chapter of the book includes a final section that provides pointers to further background and research material. Practitioners will better understand the current state of the art in data matching as well as the internal workings



160,49 €

149,99 € (zzgl. MwSt.)

Lieferfrist: bis zu 10 Tage

Artikelnummer: 9783642311635

Medium: Buch

ISBN: 978-3-642-31163-5

Verlag: Springer

Erscheinungstermin: 05.07.2012

Sprache(n): Englisch

Auflage: 2012

Serie: Data-Centric Systems and Applications

Produktform: Gebunden

Gewicht: 606 g

Seiten: 272

Format (B x H): 160 x 241 mm

