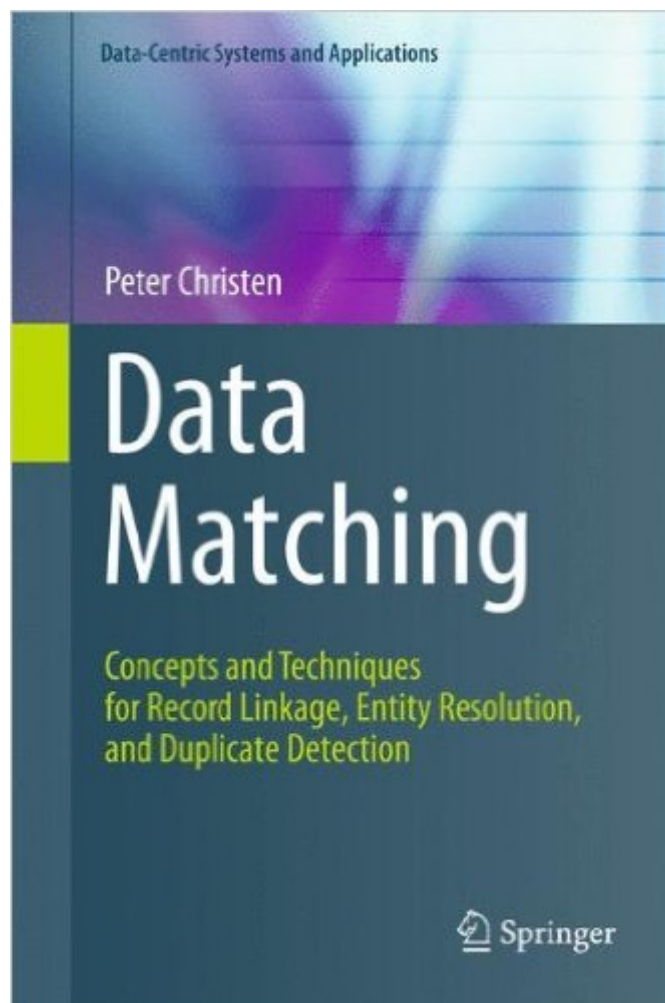


The book was found

# Data Matching: Concepts And Techniques For Record Linkage, Entity Resolution, And Duplicate Detection (Data-Centric Systems And Applications)



## Synopsis

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.

## Book Information

Series: Data-Centric Systems and Applications

Hardcover: 272 pages

Publisher: Springer; 2012 edition (July 5, 2012)

Language: English

ISBN-10: 3642311636

ISBN-13: 978-3642311635

Product Dimensions: 6.1 x 0.7 x 9.2 inches

Shipping Weight: 1.4 pounds (View shipping rates and policies)

Average Customer Review: 3.8 out of 5 stars Â Â See all reviews Â (6 customer reviews)

Best Sellers Rank: #815,449 in Books (See Top 100 in Books) #150 in Â Books > Computers & Technology > Computer Science > AI & Machine Learning > Computer Vision & Pattern Recognition #252 in Â Books > Textbooks > Computer Science > Artificial Intelligence #339 in Â Books > Computers & Technology > Networking & Cloud Computing > Network Administration > Storage & Retrieval

## Customer Reviews

I am a statistician working in a hospital. Part of my work involves data matching at a very elementary level, but I am no expert! Still, I can see that this is an important topic. The amount of data collected in health care is huge: the challenge is to match data from different sources rather than collect more data. By matching data from different sources we can gain a more complete picture of how people travel through the health care system. You can imagine the difficulties when my name is recorded as "Terence" here, "Terrance" there, "Terry" somewhere else, and, of course, "Terri" (??). Peter Christen is a computer scientist who is an expert in this field. Furthermore, he writes well. So here we have a book that is well-written, by an expert, on an important topic, published by a major publisher, at a reasonable price. Although the book is probably more advanced than I need, suprisingly, I do not regard it as too technical. There is an extensive list of references. With only about 250 pages, I can read it on the train. For a hardback book it is a good price - many university text books cost much more. "Data Matching" by Peter Christen should be in every university library.

This book is a must for anyone in the field of data linkage (or record linkage as it is also known as).It provides a comprehensive guide to all the steps in the process with clear examples. There's not too much around on this topic, so this text is much needed.

This is a fantastic book for data matching. It is written very well and easy to be understood. It strength my foundation of data matching.

[Download to continue reading...](#)

Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection (Data-Centric Systems and Applications) Detection Estimation and Modulation Theory, Part I: Detection, Estimation, and Filtering Theory Optimal Fault Detection and Resolution During

Maneuvering for Autonomous Underwater Vehicles 100 Winning Duplicate Tips (Master Bridge Series) Data Analytics: Practical Data Analysis and Statistical Guide to Transform and Evolve Any Business Leveraging the Power of Data Analytics, Data Science, ... (Hacking Freedom and Data Driven Book 2) Data Architecture: A Primer for the Data Scientist: Big Data, Data Warehouse and Data Vault Intermediate Algebra: Concepts & Applications (Bittering Concepts & Applications) Refining Design for Business: Using analytics, marketing, and technology to inform customer-centric design (Graphic Design & Visual Communication Courses) Sound Linkage (Exc Business And Economy (Whurr)) Binding and Linkage: Functional Chemistry of Biological Macromolecules Extending Simple Network Management Protocol (SNMP) Beyond Network Management: A MIB Architecture for Network-Centric Services Database Design Using Entity-Relationship Diagrams, Second Edition (Foundations of Database Design) WCF Multi-Layer Services Development with Entity Framework, 4th Edition WCF Multi-layer Services Development with Entity Framework - Fourth Edition El Super Deportista Científico / Sports Science: Anota, gana y rompe un record en tu deporte favorito/ Note, gain and break the record of your favorite sport (Spanish Edition) Goldmine Record Albums Price Guide (Goldmine Record Album Price Guide) Off the Record (The Record Series Book 1) For the Record (The Record Series Book 3) On the Record (The Record Series Book 2) Wine: The Ultimate Educational Resource Of Red Wine, Types And Origin, Red Wine Selecting & Food Pairing And Tips On Wine Occasion Matching

[Dmca](#)