
Data Matching Concepts And Techniques For Record Linkage Entity Resolution And Duplicate Detection Data Centric Systems And Applications 2012 Edition By Christen Peter Published By Springer 2012

Handbook on Impact Evaluation
Achievements and Opportunities
Data Quality
Flexible Imputation of Missing Data, Second Edition
Propensity Score Analysis
Process Analytics
Data Analytics in Medicine: Concepts, Methodologies, Tools, and Applications
Concepts and Resources for Managers
Data Mining: Concepts and Techniques
Data-Driven Policy Impact Evaluation
Decision Support Systems
The Behavioral and Social Sciences
Data Matching
Mining of Massive Datasets
Concepts, Methodologies and Techniques
Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection
Data Mining with Rattle and R
Concepts, Methodologies, Tools, and Applications
Concepts and Techniques for Querying and Analyzing Process Data
Data Matching
Database Technologies: Concepts, Methodologies, Tools, and Applications
System Engineering Analysis, Design, and Development
Registries for Evaluating Patient Outcomes
Data Quality and Record Linkage Techniques
Quantitative Methods and Practices
Research Anthology on Privatizing and Securing Data
Impact Evaluation in Practice, Second Edition
A User's Guide
Sharing Clinical Trial Data
Concepts, Strategies, and Challenges
Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection
Cognitive Analytics: Concepts, Methodologies, Tools, and Applications
Concepts, Methodologies, Tools, and Applications
Concepts, Methodologies, Tools, and Applications
Methods and Techniques for Practical Privacy-Preserving Information Sharing
Enabling Data Ecosystems for Intelligent Systems
Best Practices in Quantitative Methods
Data Mining for Business Intelligence
Concepts, Methodologies, Tools, and Applications

*Data Matching Concepts And
Techniques For Record Linkage Entity
Resolution And Duplicate Detection
Data Centric Systems And Applications
2012 Edition By Christen Peter
Published By Springer 2012*

Downloaded from db.mwpai.edu by
guest

NEAL WIGGINS

Elsevier
Data Matching Concepts and Techniques for Record Linkage,
Entity Resolution, and Duplicate Detection Springer Science &
Business Media
Handbook on Impact Evaluation National Academies Press
This book provides modern technical answers to the legal
requirements of pseudonymisation as recommended by privacy
legislation. It covers topics such as modern regulatory

frameworks for sharing and linking sensitive information, concepts and algorithms for privacy-preserving record linkage and their computational aspects, practical considerations such as dealing with dirty and missing data, as well as privacy, risk, and performance assessment measures. Existing techniques for privacy-preserving record linkage are evaluated empirically and real-world application examples that scale to population sizes are described. The book also includes pointers to freely available software tools, benchmark data sets, and tools to generate synthetic data that can be used to test and evaluate linkage techniques. This book consists of fourteen chapters grouped into four parts, and two appendices. The first part introduces the reader to the topic of linking sensitive data, the second part covers methods and techniques to link such data, the third part discusses aspects of practical importance, and the fourth part

provides an outlook of future challenges and open research problems relevant to linking sensitive databases. The appendices provide pointers and describe freely available, open-source software systems that allow the linkage of sensitive data, and provide further details about the evaluations presented. A companion Web site at <https://dmm.anu.edu.au/lstdbook2020> provides additional material and Python programs used in the book. This book is mainly written for applied scientists, researchers, and advanced practitioners in governments, industry, and universities who are concerned with developing, implementing, and deploying systems and tools to share sensitive information in administrative, commercial, or medical databases. The Book describes how linkage methods work and how to evaluate their performance. It covers all the major concepts and methods and also discusses practical matters such as computational efficiency, which are critical if the methods are to be used in practice - and it does all this in a highly accessible way! David J. Hand, Imperial College, London

Achievements and Opportunities Springer Science & Business Media

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 Quality Elsevier

Communication research is evolving and changing in a world of online journals, open-access, and new ways of obtaining data and conducting experiments via the Internet. The SAGE Encyclopedia of Communication Research Methods contains entries that cover every step of the research process, accompanied by engaging examples from the literature of communication studies. Key features include: 652 signed entries spanning four volumes, available in choice of electronic or print formats A Reader's Guide

groups entries thematically to help students interested in a specific aspect of communication research to more easily locate directly related entries Back matter includes a Chronology of the development of the field of communication research; a Resource Guide to classic books, journals, and associations; a Glossary introducing the terminology of the field; and a detailed Index Entries conclude with References/Further Readings and Cross-References to related entries to guide students further in their research journeys The Index, Reader's Guide themes, and Cross-References combine to provide robust search-and-browse in the electronic version

Flexible Imputation of Missing Data, Second Edition SAGE

This volume explores the scientific frontiers and leading edges of research across the fields of anthropology, economics, political science, psychology, sociology, history, business, education, geography, law, and psychiatry, as well as the newer, more specialized areas of artificial intelligence, child development, cognitive science, communications, demography, linguistics, and management and decision science. It includes recommendations concerning new resources, facilities, and programs that may be needed over the next several years to ensure rapid progress and provide a high level of returns to basic research.

Propensity Score Analysis Springer Science & Business Media

"This reference expands the field of database technologies through four-volumes of in-depth, advanced research articles from nearly 300 of the world's leading professionals"--Provided by publisher.

Process Analytics IGI Global

Missing data pose challenges to real-life data analysis. Simple ad-hoc fixes, like deletion or mean imputation, only work under highly restrictive conditions, which are often not met in practice. Multiple imputation replaces each missing value by multiple plausible values. The variability between these replacements reflects our ignorance of the true (but missing) value. Each of the completed data set is then analyzed by standard methods, and the results are pooled to obtain unbiased estimates with correct confidence intervals. Multiple imputation is a general approach that also inspires novel solutions to old problems by reformulating the task at hand as a missing-data problem. This is the second edition of a popular book on multiple imputation, focused on explaining the application of methods through detailed worked examples using the MICE package as developed by the author. This new edition incorporates the recent developments in this fast-moving field. This class-tested book avoids mathematical and technical details as much as possible: formulas are accompanied by verbal statements that explain the formula in accessible terms. The book sharpens the reader's intuition on how to think about missing data, and provides all the tools needed to execute a well-grounded quantitative analysis in the presence of missing data.

Data Analytics in Medicine: Concepts, Methodologies, Tools, and Applications CRC Press

With the proliferation of huge amounts of (heterogeneous) data on the Web, the importance of information retrieval (IR) has grown considerably over the last few years. Big players in the computer industry, such as Google, Microsoft and Yahoo!, are the primary contributors of technology for fast access to Web-based information; and searching capabilities are now integrated into most information systems, ranging from business management software and customer relationship systems to social networks and mobile phone applications. Ceri and his co-authors aim at taking their readers from the foundations of modern information retrieval to the most advanced challenges of Web IR. To this end, their book is divided into three parts. The first part addresses the principles of IR and provides a systematic and compact

description of basic information retrieval techniques (including binary, vector space and probabilistic models as well as natural language search processing) before focusing on its application to the Web. Part two addresses the foundational aspects of Web IR by discussing the general architecture of search engines (with a focus on the crawling and indexing processes), describing link analysis methods (specifically Page Rank and HITS), addressing recommendation and diversification, and finally presenting advertising in search (the main source of revenues for search engines). The third and final part describes advanced aspects of Web search, each chapter providing a self-contained, up-to-date survey on current Web research directions. Topics in this part include meta-search and multi-domain search, semantic search, search in the context of multimedia data, and crowd search. The book is ideally suited to courses on information retrieval, as it covers all Web-independent foundational aspects. Its presentation is self-contained and does not require prior background knowledge. It can also be used in the context of classic courses on data management, allowing the instructor to cover both structured and unstructured data in various formats. Its classroom use is facilitated by a set of slides, which can be downloaded from www.search-computing.org.

Concepts and Resources for Managers SAGE

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 Mining: Concepts and Techniques Elsevier

With the immense amount of data that is now available online, security concerns have been an issue from the start, and have grown as new technologies are increasingly integrated in data collection, storage, and transmission. Online cyber threats, cyber terrorism, hacking, and other cybercrimes have begun to take

advantage of this information that can be easily accessed if not properly handled. New privacy and security measures have been developed to address this cause for concern and have become an essential area of research within the past few years and into the foreseeable future. The ways in which data is secured and privatized should be discussed in terms of the technologies being used, the methods and models for security that have been developed, and the ways in which risks can be detected, analyzed, and mitigated. The Research Anthology on Privatizing and Securing Data reveals the latest tools and technologies for privatizing and securing data across different technologies and industries. It takes a deeper dive into both risk detection and mitigation, including an analysis of cybercrimes and cyber threats, along with a sharper focus on the technologies and methods being actively implemented and utilized to secure data online. Highlighted topics include information governance and privacy, cybersecurity, data protection, challenges in big data, security threats, and more. This book is essential for data analysts, cybersecurity professionals, data scientists, security analysts, IT specialists, practitioners, researchers, academicians, and students interested in the latest trends and technologies for privatizing and securing data.

Data-Driven Policy Impact Evaluation Springer Science & Business Media

This open access book explores the dataspace paradigm as a best-effort approach to data management within data ecosystems. It establishes the theoretical foundations and principles of real-time linked dataspaces as a data platform for intelligent systems. The book introduces a set of specialized best-effort techniques and models to enable loose administrative proximity and semantic integration for managing and processing events and streams. The book is divided into five major parts: Part I "Fundamentals and Concepts" details the motivation behind and core concepts of real-time linked dataspaces, and establishes the need to evolve data management techniques in order to meet the challenges of enabling data ecosystems for intelligent systems within smart environments. Further, it explains the fundamental concepts of dataspaces and the need for specialization in the processing of dynamic real-time data. Part II "Data Support Services" explores the design and evaluation of critical services, including catalog, entity management, query and search, data service discovery, and human-in-the-loop. In turn, Part III "Stream and Event Processing Services" addresses the design and evaluation of the specialized techniques created for real-time support services including complex event processing, event service composition, stream dissemination, stream matching, and approximate semantic matching. Part IV "Intelligent Systems and Applications" explores the use of real-time linked dataspaces within real-world smart environments. In closing, Part V "Future Directions" outlines future research challenges for dataspaces, data ecosystems, and intelligent systems. Readers will gain a detailed understanding of how the dataspace paradigm is now being used to enable data ecosystems for intelligent systems within smart environments. The book covers the fundamental theory, the creation of new techniques needed for support services, and lessons learned from real-world intelligent systems and applications focused on sustainability. Accordingly, it will benefit not only researchers and graduate students in the fields of data management, big data, and IoT, but also professionals who need to create advanced data management platforms for intelligent systems, smart environments, and data ecosystems.

Decision Support Systems John Wiley & Sons

Advancements in data science have created opportunities to sort, manage, and analyze large amounts of data more effectively and

efficiently. Applying these new technologies to the healthcare industry, which has vast quantities of patient and medical data and is increasingly becoming more data-reliant, is crucial for refining medical practices and patient care. *Data Analytics in Medicine: Concepts, Methodologies, Tools, and Applications* is a vital reference source that examines practical applications of healthcare analytics for improved patient care, resource allocation, and medical performance, as well as for diagnosing, predicting, and identifying at-risk populations. Highlighting a range of topics such as data security and privacy, health informatics, and predictive analytics, this multi-volume book is ideally designed for doctors, hospital administrators, nurses, medical professionals, IT specialists, computer engineers, information technologists, biomedical engineers, data-processing specialists, healthcare practitioners, academicians, and researchers interested in current research on the connections between data analytics in the field of medicine.

The Behavioral and Social Sciences SAGE

How do you approach answering queries when your data is stored in multiple databases that were designed independently by different people? This is first comprehensive book on data integration and is written by three of the most respected experts in the field. This book provides an extensive introduction to the theory and concepts underlying today's data integration techniques, with detailed, instruction for their application using concrete examples throughout to explain the concepts. Data integration is the problem of answering queries that span multiple data sources (e.g., databases, web pages). Data integration problems surface in multiple contexts, including enterprise information integration, query processing on the Web, coordination between government agencies and collaboration between scientists. In some cases, data integration is the key bottleneck to making progress in a field. The authors provide a working knowledge of data integration concepts and techniques, giving you the tools you need to develop a complete and concise package of algorithms and applications. Offers a range of data integration solutions enabling you to focus on what is most relevant to the problem at hand Enables you to build your own algorithms and implement your own data integration applications

Data Matching National Academies Press

This book provides step-by-step instructions on how to analyze text generated from in-depth interviews and focus groups, relating predominantly to applied qualitative studies. The book covers all aspects of the qualitative data analysis process, employing a phenomenological approach which has a primary aim of describing the experiences and perceptions of research participants. Similar to Grounded Theory, the authors' approach is inductive, content-driven, and searches for themes within textual data.

Mining of Massive Datasets World Bank Publications

The second edition of the *Impact Evaluation in Practice* handbook is a comprehensive and accessible introduction to impact evaluation for policy makers and development practitioners. First published in 2011, it has been used widely across the development and academic communities. The book incorporates real-world examples to present practical guidelines for designing and implementing impact evaluations. Readers will gain an understanding of impact evaluations and the best ways to use them to design evidence-based policies and programs. The updated version covers the newest techniques for evaluating programs and includes state-of-the-art implementation advice, as well as an expanded set of examples and case studies that draw on recent development challenges. It also includes new material on research ethics and partnerships to conduct impact evaluation. The handbook is divided into four sections: Part One

discusses what to evaluate and why; Part Two presents the main impact evaluation methods; Part Three addresses how to manage impact evaluations; Part Four reviews impact evaluation sampling and data collection. Case studies illustrate different applications of impact evaluations. The book links to complementary instructional material available online, including an applied case as well as questions and answers. The updated second edition will be a valuable resource for the international development community, universities, and policy makers looking to build better evidence around what works in development.

Concepts, Methodologies and Techniques John Wiley & Sons

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.

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

Data mining is the art and science of intelligent data analysis. By building knowledge from information, data mining adds considerable value to the ever increasing stores of electronic data that abound today. In performing data mining many decisions need to be made regarding the choice of methodology, the choice of data, the choice of tools, and the choice of algorithms. Throughout this book the reader is introduced to the basic concepts and some of the more popular algorithms of data mining. With a focus on the hands-on end-to-end process for data mining, Williams guides the reader through various capabilities of the easy to use, free, and open source Rattle Data Mining Software built on the sophisticated R Statistical Software. The focus on doing data mining rather than just reading about data mining is refreshing. The book covers data understanding, data preparation, data refinement, model building, model evaluation, and practical deployment. The reader will learn to rapidly deliver

a data mining project using software easily installed for free from the Internet. Coupling Rattle with R delivers a very sophisticated data mining environment with all the power, and more, of the many commercial offerings.

Data Mining with Rattle and R IGI Global

With the increased use of technology in modern society, high volumes of multimedia information exists. It is important for businesses, organizations, and individuals to understand how to optimize this data and new methods are emerging for more efficient information management and retrieval. Information Retrieval and Management: Concepts, Methodologies, Tools, and Applications is an innovative reference source for the latest academic material in the field of information and communication technologies and explores how complex information systems interact with and affect one another. Highlighting a range of topics such as knowledge discovery, semantic web, and information resources management, this multi-volume book is ideally designed for researchers, developers, managers, strategic planners, and advanced-level students.

Concepts, Methodologies, Tools, and Applications Government Printing Office

Fully updated to reflect the most recent changes in the field, the Second Edition of Propensity Score Analysis provides an accessible, systematic review of the origins, history, and statistical foundations of propensity score analysis, illustrating how it can be used for solving evaluation and causal-inference problems. With a strong focus on practical applications, the authors explore various strategies for employing PSA, discuss the use of PSA with alternative types of data, and delineate the limitations of PSA under a variety of constraints. Unlike existing textbooks on program evaluation and causal inference, this book delves into statistical concepts, formulas, and models within the context of a robust and engaging focus on application.

Concepts and Techniques for Querying and Analyzing Process Data Greenwood Publishing Group

Public programs are designed to reach certain goals and beneficiaries. Methods to understand whether such programs actually work, as well as the level and nature of impacts on intended beneficiaries, are main themes of this book.

Best Sellers - Books :

- [A Letter From Your Teacher: On The First Day Of School](#)
- [Bluey And Bingo's Fancy Restaurant Cookbook: Yummy Recipes, For Real Life By Penguin Young Readers Licenses](#)
- [Twisted Love \(twisted, 1\)](#)
- [World Of Eric Carle, Around The Farm 30-button Animal Sound Book - Great For First Words - Pi Kids By Pi Kids](#)
- [What To Expect When You're Expecting By Heidi Murkoff](#)
- [Bluey And Bingo's Fancy Restaurant Cookbook: Yummy Recipes, For Real Life](#)
- [The Shadow Work Journal: A Guide To Integrate And Transcend Your Shadows By Keila Shaheen](#)
- [A Court Of Wings And Ruin \(a Court Of Thorns And Roses, 3\) By Sarah J. Maas](#)
- [8 Rules Of Love: How To Find It, Keep It, And Let It Go](#)
- [If Animals Kissed Good Night By Ann Whitford Paul](#)