Data Matching Concepts And Techniques For Record Linkage Entity Resolution And Duplicate Detection Data Centric Systems And Applications

As recognized, adventure as skillfully as experience practically lesson, amusement, as competently as promise can be gotten by just checking out a ebook **Data Matching Concepts And Techniques For Record Linkage Entity Resolution And Duplicate Detection Data Centric Systems And Applications** plus it is not directly done, you could acknowledge even more approaching this life, in the region of the world.

We offer you this proper as with ease as easy habit to acquire those all. We offer Data Matching Concepts And Techniques For Record Linkage Entity Resolution And Duplicate Detection Data Centric Systems And Applications and numerous book collections from fictions to scientific research in any way. along with them is this Data Matching Concepts And Techniques For Record Linkage Entity Resolution And Duplicate Detection Data Centric Systems And Applications that can be your partner.

Enabling Data Ecosystems for Intelligent Systems SAGE

```
This User's Guide is intended to support the
design, implementation, analysis,
interpretation, and quality evaluation of
registries created to increase understanding
of patient outcomes. For the purposes of this
guide, a patient registry is an organized
system that uses observational study methods
to collect uniform data (clinical and other)
to evaluate specified outcomes for a
population defined by a particular disease,
condition, or exposure, and that serves one or
more predetermined scientific, clinical, or
policy purposes. A registry database is a file
(or files) derived from the registry. Although
registries can serve many purposes, this guide
focuses on registries created for one or more
of the following purposes: to describe the
natural history of disease, to determine
clinical effectiveness or cost-effectiveness
of health care products and services, to
measure or monitor safety and harm, and/or to
measure quality of care. Registries are
classified according to how their populations
are defined. For example, product registries
include patients who have been exposed to
biopharmaceutical products or medical devices.
Health services registries consist of patients
who have had a common procedure, clinical
encounter, or hospitalization. Disease or
condition registries are defined by patients
having the same diagnosis, such as cystic
fibrosis or heart failure. The User's Guide
was created by researchers affiliated with
AHRQ's Effective Health Care Program,
particularly those who participated in AHRQ's
DEcIDE (Developing Evidence to Inform
Decisions About Effectiveness) program.
Chapters were subject to multiple internal and
external independent reviews.
```

Best Practices in Quantitative Methods Elsevier

Due to the growing use of web applications and communication devices, the use of data has increased throughout various industries, including business and healthcare. It is necessary to develop specific software programs that can analyze and interpret large amounts of data quickly in order to ensure adequate usage and predictive results. Cognitive Analytics: Concepts, Methodologies, Tools, and Applications provides emerging perspectives on the theoretical and practical aspects of data analysis tools and techniques. It also examines the incorporation of pattern management as well as decision-making and prediction processes through the use of data management and analysis. Highlighting a range of topics such as natural language processing, big data, and pattern recognition, this multi-volume book is ideally designed for information technology professionals, software developers, data analysts, graduate-level students, researchers, computer engineers, software engineers, IT specialists, and academicians.

Process Analytics Springer Nature

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.

Real-time Linked Dataspaces Elsevier

Data Mining: Concepts and Techniques provides the concepts and techniques in processing gathered data or information, which will be used in various applications. Specifically, it explains data mining and the tools used in discovering knowledge from the collected data. This book is referred as the knowledge discovery from data (KDD). It focuses on the feasibility, usefulness, effectiveness, and scalability of techniques of large data sets. After describing data mining, this edition explains the methods of knowing, preprocessing, processing, and warehousing data. It then presents information about data warehouses, online analytical processing (OLAP), and data cube technology. Then, the methods involved in mining frequent patterns, associations, and correlations for large data sets are described. The book details the methods for data classification and introduces the concepts and methods for data clustering. The remaining chapters discuss

the outlier detection and the trends, applications, and research frontiers in data mining. This book is intended for Computer Science students, application developers, business professionals, and researchers who seek information on data mining. Presents dozens of algorithms and implementation examples, all in pseudo-code and suitable for use in real-world, large-scale data mining projects Addresses advanced topics such as mining object-relational databases, spatial databases, multimedia databases, time-series databases, text databases, the World Wide Web, and applications in several fields Provides a comprehensive, practical look at the concepts and techniques you need to get the most out of your data

Impact Evaluation in Practice, Second Edition Springer Nature

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.

Linking Sensitive Data IGI Global

Now in its second edition, this book focuses on practical algorithms for mining data from even the largest datasets.

Methods and Techniques for Practical Privacy-Preserving Information Sharing John Wiley & Sons

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

Research Anthology on Privatizing and Securing Data John Wiley & Sons

Security and authentication issues are surging to the forefront of the research realm in global society. As technology continues to evolve, individuals are finding it easier to infiltrate various forums and facilities where they can illegally obtain information and access. By implementing biometric authentications to these forums, users are able to prevent attacks on their privacy and security. Biometrics: Concepts, Methodologies, Tools, and Applications is a multi-volume publication highlighting critical topics related to access control, user identification, and surveillance technologies. Featuring emergent research on the issues and challenges in security and privacy, various forms of user authentication, biometric applications to image processing and computer vision, and security applications within the field, this publication is an ideal reference source for researchers, engineers, technology developers, students, and security specialists.

How Access to Microdata is Transforming Policy Design 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.

Quantitative Methods and Practices Elsevier

Data sharing can accelerate new discoveries by avoiding duplicative trials, stimulating new ideas for research, and enabling the maximal scientific knowledge and benefits to be gained from the efforts of clinical trial participants and investigators. At the same time, sharing clinical trial data presents risks, burdens, and challenges. These include the need to protect the privacy and honor the consent of clinical trial participants; safeguard the legitimate economic interests of sponsors; and guard against invalid secondary analyses, which could undermine trust in clinical trials or otherwise harm public health. Sharing Clinical Trial Data presents activities and strategies for the responsible sharing of clinical trial data. With the goal of increasing scientific knowledge to lead to better therapies for patients, this book identifies guiding principles and makes recommendations to maximize the benefits and minimize risks. This report offers guidance on the types of clinical trial data available at different points in the process, the points in the process at which each type of data should be shared, methods for sharing data, what groups should have access to data, and future knowledge and infrastructure needs. Responsible sharing of clinical trial data will allow other investigators to replicate published findings and carry out additional analyses, strengthen the evidence base for regulatory and clinical decisions, and increase the scientific knowledge gained from investments by the funders of clinical trials. The recommendations of Sharing Clinical Trial Data will be useful both now and well into the future as improved sharing of data leads to a stronger evidence base for treatment. This book will be of interest to stakeholders across the spectrum of research--from funders, to researchers, to journals, to physicians, and ultimately, to patients.

Data Matching Springer Science & Business Media

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

Page 1/2

Data Matching Concepts And Techniques For Record Linkage Entity Resolution And Duplicate Detection Data Centric Systems And Applications

April, 27 2024

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.

Applied Thematic Analysis Academic Press
Learn how to develop models for classification, prediction, and customer segmentation with the help of Data Mining for Business Intelligence In today's world, businesses are becoming more capable of accessing their ideal consumers, and an understanding of data mining contributes to this success. Data Mining for Business Intelligence, which was developed from a course taught at the Massachusetts Institute of Technology's Sloan School of Management, and the University of Maryland's Smith School of Business, uses real data and actual cases to illustrate the applicability of data mining intelligence to the development of successful business models. Featuring XLMiner, the Microsoft Office Excel add-in, this book allows readers to follow along and implement algorithms at their own speed, with a minimal learning curve. In addition, students and practitioners of data mining techniques are presented with hands-on, business-oriented applications. An abundant amount of exercises and examples are provided to motivate learning and understanding. Data Mining for Business Intelligence: Provides both a theoretical and practical understanding of the key methods of classification, prediction, reduction, exploration, and affinity analysis Features a business decision-making context for these key methods Illustrates the application and interpretation of these methods using real business cases and data This book helps readers understand the beneficial relationship that can be established between data mining and smart business practices, and is an excellent learning tool for creating valuable strategies and making wiser business decisions.

Encyclopedia of Research Design Springer
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/lsdbook2020 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

Data-Driven Policy Impact Evaluation Springer
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.

Data Mining for Business Intelligence IGI Global
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.

Concepts and Techniques for Querying and Analyzing Process Data National Academies Press
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.

Handbook on Impact Evaluation Data MatchingConcepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection
In the light of better and more detailed administrative databases, this open access book provides statistical tools for evaluating the effects of public policies advocated by governments and public institutions. Experts from academia, national statistics offices and various research centers present modern econometric methods for an efficient data-driven policy evaluation and monitoring, assess the causal effects of policy measures and report on best practices of successful data management and usage. Topics include data confidentiality, data linkage, and national practices in policy areas such as public health, education and employment. It offers scholars as well as practitioners from public administrations, consultancy firms and nongovernmental organizations insights into

counterfactual impact evaluation methods and the potential of data-based policy and program evaluation.

Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner Springer Science & Business Media
Entity Resolution and Information Quality presents topics and definitions, and clarifies confusing terminologies regarding entity resolution and information quality. It takes a very wide view of IQ, including its six-domain framework and the skills formed by the International Association for Information and Data Quality {IAIDQ}. The book includes chapters that cover the principles of entity resolution and the principles of Information Quality, in addition to their concepts and terminology. It also discusses the Fellegi-Sunter theory of record linkage, the Stanford Entity Resolution Framework, and the Algebraic Model for Entity Resolution, which are the major theoretical models that support Entity Resolution. In relation to this, the book briefly discusses entity-based data integration (EBDI) and its model, which serve as an extension of the Algebraic Model for Entity Resolution. There is also an explanation of how the three commercial ER systems operate and a description of the non-commercial open-source system known as OYSTER. The book concludes by discussing trends in entity resolution research and practice. Students taking IT courses and IT professionals will find this book invaluable. First authoritative reference explaining entity resolution and how to use it effectively Provides practical system design advice to help you get a competitive advantage Includes a companion site with synthetic customer data for applicatory exercises, and access to a Java-based Entity Resolution program.

Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection Springer
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, Methodologies, Tools, and Applications IGI Global
"Comprising more than 500 entries, the Encyclopedia of Research Design explains how to make decisions about research design, undertake research projects in an ethical manner, interpret and draw valid inferences from data, and evaluate experiment design strategies and results. Two additional features carry this encyclopedia far above other works in the field: bibliographic entries devoted to significant articles in the history of research design and reviews of contemporary tools, such as software and statistical procedures, used to analyze results. It covers the spectrum of research design strategies, from material presented in introductory classes to topics necessary in graduate research; it addresses cross- and multidisciplinary research needs, with many examples drawn from the social and behavioral sciences, neurosciences, and biomedical and life sciences; it provides summaries of advantages and disadvantages of often-used strategies; and it uses hundreds of sample tables, figures, and equations based on real-life cases."--Publisher's description.

Page 2/2

Data Matching Concepts And Techniques For Record Linkage Entity Resolution And Duplicate Detection Data Centric Systems And Applications

April, 27 2024