Foundations Of Data Quality Management Synthesis Lectures On Data Man Pdf Pdf

Foundations Of Data Quality Management Synthesis Lectures On Data Man Pdf Pdf - As recognized, adventure as well as experience just about lesson, amusement, as well as concurrence can be gotten by just checking out a book **foundations of data quality management synthesis lectures on data man pdf pdf** afterward it is not directly done, you could assume even more on this life, approaching the world.

We present you this proper as with ease as simple habit to get those all. We present foundations of data quality management synthesis lectures on data man pdf pdf and numerous book collections from fictions to scientific research in any way. among them is this foundations of data quality management synthesis lectures on data man pdf pdf that can be your partner. Yeah, reviewing a books **foundations of data quality management synthesis lectures on data man pdf pdf** could grow your close friends listings. This is just one of the solutions for you to be successful. As understood, capability does not recommend that you have astounding points.

Comprehending as skillfully as accord even more than further will find the money for each success. next to, the broadcast as well as perspicacity of this foundations of data quality management synthesis lectures on data man pdf pdf can be taken as well as picked to act. - *Foundations Of Data Quality Management Synthesis Lectures On Data Man Pdf Pdf*

Foundations Of Data Quality Management Synthesis Lectures On Data Man Pdf Pdf (2023)

Data Cleaning Ihab F. Ilyas 2019-06-18 Data quality is one of the most important problems in data management, since dirty data often leads to inaccurate data analytics results and incorrect business decisions. Poor data across businesses and the U.S. government are reported to cost trillions of dollars a year. Multiple surveys show that dirty data is the most common barrier faced by data scientists. Not surprisingly, developing effective and efficient data cleaning solutions is challenging and is rife with deep theoretical and engineering problems. This book is about data cleaning, which is used to refer to all kinds of tasks and activities to detect and repair errors in the data. Rather than focus on a particular data cleaning task, we give an overview of the end-to-end data cleaning process, describing various error detection and repair methods, and attempt to anchor these proposals with multiple taxonomies and views. Specifically, we cover four of the most common and important data cleaning tasks, namely, outlier detection, data transformation, error repair (including imputing missing values), and data deduplication. Furthermore, due to the increasing popularity and applicability of machine learning techniques, we include a chapter that specifically explores how machine learning techniques are used for data cleaning, and how data cleaning is used to improve machine learning models. This book is intended to serve as a useful reference for researchers and practitioners who are interested in the area of data quality and data cleaning. It can also be used as a textbook for a graduate course. Although we aim at covering state-of-the-art algorithms and techniques, we recognize that data cleaning is still an active field of research and therefore provide future directions of research whenever appropriate.
Flexible Query Answering Systems Alfredo Cuzzocrea 2019-09-11 This book constitutes the refereed proceedings of the 13th International Conference on Flexible Query Answering Systems, FQAS 2019, held in Amantea, Italy, in July 2019. The 27 full papers and 10 short papers presented were carefully reviewed and selected from 43 submissions. The papers present emerging research trends with a special focus on flexible querying and analytics for smart cities and smart societies in the age of big data. They are organized in the following topical sections: flexible database management and querying; ontologies and knowledge bases; social networks and social media; argumentation-based query answering; data mining and knowledge discovery; advanced flexible query answering methodologies and techniques; flexible query answering methods and techniques; flexible intelligent information-oriented and network-oriented approaches; big data veracity and soft computing; flexibility in tools; and systems and miscellanea.
Computational Intelligence, Cyber Security and Computational Models G. Sai Sundara Krishnan 2013-11-26 This book contains cutting-edge research material presented by researchers, engineers, developers, and practitioners from academia and industry at the International Conference on Computational Intelligence, Cyber Security and Computational Models (ICC3) organized by PSG College of Technology, Coimbatore, India during December 19–21, 2013. The materials in the book include theory and applications to provide design, analysis, and modeling of the key areas. The book will be useful material for students, researchers, professionals, as well academicians in understanding current research trends and findings and future scope of research in computational intelligence, cyber security, and computational models.
Data Management in the Cloud Divyakant Agrawal 2022-05-31 Cloud computing has emerged as a successful paradigm of service-oriented computing and has revolutionized the way computing infrastructure is used. This success has seen a proliferation in the number of applications that are being deployed in various cloud platforms. There has also been an increase in the scale of the data generated as well as consumed by such applications. Scalable database management systems form a critical part of the cloud infrastructure. The attempt to address the challenges posed by the management of big data has led to a plethora of systems. This book aims to clarify some of the important concepts in the design space of scalable data management in cloud computing infrastructures. Some of the questions that this book aims to answer are: the appropriate systems for a specific set of application requirements, the research challenges in data management for the cloud, and what is novel in the cloud for database researchers? We also aim to address one basic question: whether cloud computing poses new challenges in scalable data management or it is just a reincarnation of old problems? We provide a comprehensive background study of state-of-the-art systems for scalable data management and analysis. We also identify important aspects in the design of different systems and the applicability and scope of these systems. A thorough understanding of current solutions and a precise characterization of the design space are essential for clearing the "cloudy skies of data management" and ensuring the success of DBMSs in the cloud, thus emulating the success enjoyed by relational databases in traditional enterprise settings. Table of Contents: Introduction / Distributed Data Management / Cloud Data Management: Early Trends / Transactions on Co-located Data / Transactions on Distributed Data / Multi-tenant Database Systems / Concluding Remarks
Veracity of Data Laure Berti-Équille 2022-05-31 On the Web, a massive amount of user-generated content is available through various channels (e.g., texts, tweets, Web tables, databases, multimedia-sharing platforms, etc.). Conflicting information, rumors, erroneous and fake content can be easily spread across multiple sources, making it hard to distinguish between what is true and what is not. This book gives an overview of fundamental issues and recent contributions for ascertaining the veracity of data in the era of Big Data. The text is organized into six chapters, focusing on structured data extracted from texts. Chapter 1 introduces the problem of ascertaining the veracity of data in a multi-source and evolving context. Issues related to information extraction are presented in Chapter 2. Current truth discovery computation algorithms are presented in details in Chapter 3. It is followed by practical techniques for evaluating data source reputation and authoritativeness in Chapter 4. The theoretical foundations and various approaches for modeling diffusion phenomenon of misinformation spreading in networked systems are studied in Chapter 5. Finally, truth discovery computation from extracted data in a dynamic context of misinformation propagation raises interesting challenges that are explored in Chapter 6. This text is intended for a seminar course at the graduate level. It is also to serve as a useful resource for researchers and practitioners who are interested in the study of fact-checking, truth discovery, or rumor spreading.
Foundations of Information and Knowledge Systems Thomas Lukasiewicz 2012-02-29 This book constitutes the proceedings of the 7th International Symposium on Foundations of Information and Knowledge Systems, FoIKS 2012, held in Kiel, Germany, in March 2012. The 12 regular and 8 short papers, presented together with two invited talks in full paper-length, were carefully reviewed and selected from 53 submissions. The contributions cover foundational aspects of information and knowledge systems. These include the application of ideas, theories or methods from specific disciplines to information and knowledge systems, such as discrete mathematics, logic and algebra, model theory, informaiton theory, complexity theory, algorithmics and computation, statistics, and optimization.
Web Technologies and Applications Lei Chen 2014-08-15 This book constitutes the refereed proceedings of the 16th Asia-Pacific Conference APWeb 2014 held in Changsha, China, in September 2014. The 34 full papers and 23 short papers presented were carefully reviewed and selected from 134 submissions. The papers address research, development and advanced applications of large-scale data management, web and search technologies, and information processing.
[illegible] [illegible] 2018-05-01 [illegible]“[illegible]”[illegible]cyberspace[illegible]
Modern Management Based on Big Data II and Machine Learning and Intelligent Systems III A.J. Tallón-Ballesteros 2021-12-03 It is data that guides the path of applications, and Big Data technologies are enabling new paths which can deal with information in a reasonable time to arrive at an approximate solution, rather than a more exact result in an unacceptably long time. This can be particularly important

when dealing with an urgent issue such as that of the COVID-19 pandemic. This book presents the proceedings of two conferences: MMBD 2021 and MLIS 2021. The MMBD conference deals with two main subjects; those of Big Data and Modern Management. The MLIS conference aims to provide a platform for knowledge exchange of the most recent scientific and technological advances in the field of machine learning and intelligent systems. Both conferences were originally scheduled to be held from 8-11 November 2021, in Quanzhou, China and Xiamen, China respectively. Both conferences were ultimately held fully online on the same dates, hosted by Huaqiao University in Quanzhou and Xiamen respectively. The book is in two parts, and contains a total of 78 papers (54 from MMBD2021 and 24 from MLIS2021) selected after rigorous review from a total of some 300 submissions. The reviewers bore in mind the breadth and depth of the research topics that fall within the scope of MMBD and MLIS, and selected the 78 most promising and FAIA mainstream-relevant contributions for inclusion in this two-part volume. All the papers present original ideas or results of general significance supported by clear reasoning, compelling evidence and rigorous methods.

Data Processing on FPGAs Jens Teubner 2022-05-31 Roughly a decade ago, power consumption and heat dissipation concerns forced the semiconductor industry to radically change its course, shifting from sequential to parallel computing. Unfortunately, improving performance of applications has now become much more difficult than in the good old days of frequency scaling. This is also affecting databases and data processing applications in general, and has led to the popularity of so-called data appliances—specialized data processing engines, where software and hardware are sold together in a closed box. Field-programmable gate arrays (FPGAs) increasingly play an important role in such systems. FPGAs are attractive because the performance gains of specialized hardware can be significant, while power consumption is much less than that of commodity processors. On the other hand, FPGAs are way more flexible than hard-wired circuits (ASICs) and can be integrated into complex systems in many different ways, e.g., directly in the network for a high-frequency trading application. This book gives an introduction to FPGA technology targeted at a database audience. In the first few chapters, we explain in detail the inner workings of FPGAs. Then we discuss techniques and design patterns that help mapping algorithms to FPGA hardware so that the inherent parallelism of these devices can be leveraged in an optimal way. Finally, the book will illustrate a number of concrete examples that exploit different advantages of FPGAs for data processing. Table of Contents: Preface / Introduction / A Primer in Hardware Design / FPGAs / FPGA Programming Models / Data Stream Processing / Accelerated DB Operators / Secure Data Processing / Conclusions / Bibliography / Authors' Biographies / Index

Integrity Constraints on Rich Data Types Shaoxu Song 2023-03-30 This book examines the recent trend of extending data dependencies to adapt to rich data types in order to address variety and veracity issues in big data. Readers will be guided through the full range of rich data types where data dependencies have been successfully applied, including categorical data with equality relationships, heterogeneous data with similarity relationships, numerical data with order relationships, sequential data with timestamps, and graph data with complicated structures. The text will also discuss interesting constraints on ordering or similarity relationships contained in novel classes of data dependencies in addition to those in equality relationships, e.g., considered in functional dependencies (FDs). In addition to exploring the concepts of these data dependency notations, the book investigates the extension relationships between data dependencies, such as conditional functional dependencies (CFDs) that extend conventional functional dependencies (FDs). This forms in the book a family tree of extensions, mostly rooted in FDs, that help illuminate the expressive power of various data dependencies. Moreover, the book points to work on the discovery of dependencies from data, since data dependencies are often unlikely to be manually specified in a traditional way, given the huge volume and high variety in big data. It further outlines the applications of the extended data dependencies, in particular in data quality practice. Altogether, this book provides a comprehensive guide for readers to select proper data dependencies for their applications that have sufficient expressive power and reasonable discovery cost. Finally, the book concludes with several directions of future studies on emerging data.

Linked Data Tom Heath 2011 The World Wide Web has enabled the creation of a global information space comprising linked documents. As the Web becomes ever more enmeshed with our daily lives, there is a growing desire for direct access to raw data not currently available on the Web or bound up in hypertext documents. Linked Data provides a publishing paradigm in which not only documents, but also data, can be a first class citizen of the Web, thereby enabling the extension of the Web with a global data space based on open standards - the Web of Data. In this Synthesis lecture we provide readers with a detailed technical introduction to Linked Data. We begin by outlining the basic principles of Linked Data, including coverage of relevant aspects of Web architecture. The remainder of the text is based around two main themes - the publication and consumption of Linked Data. Drawing on a practical Linked Data scenario, we provide guidance and best practices on: architectural approaches to publishing Linked Data; choosing URIs and vocabularies to identify and describe resources; deciding what data to return in a description of a resource on the Web; methods and frameworks for automated linking of data sets; and testing and debugging approaches for Linked Data deployments. We give an overview of existing Linked Data applications and then examine the architectures that are used to consume Linked Data from the Web, alongside existing tools and frameworks that enable these. Readers can expect to gain a rich technical understanding of Linked Data fundamentals, as the basis for application development, research or further study. Table of Contents: List of Figures / Introduction / Principles of Linked Data / The Web of Data / Linked Data Design Considerations / Recipes for Publishing Linked Data / Consuming Linked Data / Summary and Outlook

Perspectives on Business Intelligence Raymond T. Ng 2022-05-31 In the 1980s, traditional Business Intelligence (BI) systems focused on the delivery of reports that describe the state of business activities in the past, such as for questions like "How did our sales perform during the last quarter?" A decade later, there was a shift to more interactive content that presented how the business was performing at the present time, answering questions like "How are we doing right now?" Today the focus of BI users are looking into the future. "Given what I did before and how I am currently doing this quarter, how will I do next quarter?" Furthermore, fuelled by the demands of Big Data, BI systems are going through a time of incredible change. Predictive analytics, high volume data, unstructured data, social data, mobile, consumable analytics, and data visualization are all examples of demands and capabilities that have become critical within just the past few years, and are growing at an unprecedented pace. This book introduces research problems and solutions on various aspects central to next-generation BI systems. It begins with a chapter on an industry perspective on how BI has evolved, and discusses how game-changing trends have drastically reshaped the landscape of BI. One of the game changers is the shift toward the consumerization of BI tools. As a result, for BI tools to be successfully used by business users (rather than IT departments), the tools need a business model, rather than a data model. One chapter of the book surveys four different types of business modeling. However, even with the existence of a business model for users to express queries, the data that can meet the needs are still captured within a data model. The next chapter on vivification addresses the problem of closing the gap, which is often significant, between the business and the data models. Moreover, Big Data forces BI systems to integrate and consolidate multiple, and often wildly different, data sources. One chapter gives an overview of several integration architectures for dealing with the challenges that need to be overcome. While the book so far focuses on the usual structured relational data, the remaining chapters turn to unstructured data, an ever-increasing and important component of Big Data. One chapter on information extraction describes methods for dealing with the extraction of relations from free text and the web. Finally, BI users need tools to visualize and interpret new and complex types of information in a way that is compelling, intuitive, but accurate. The last chapter gives an overview of information visualization for decision support and text.

Large-Scale Agile Frameworks Sascha Block 2023-08-17 The book Large-Scale Agile Frameworks provides practical solutions for cross-team and cross-functional prioritization of requirements and documentation for enterprises. It reflects the interplay of current technology trends such as cloud computing and organizational requirements for microservices. Organizations are increasingly required to align their IT strategy with customer needs for customer-centric and service-oriented products and services. The book analyzes the unique requirements of a differentiated software service offering and shows how agile principles are effective in addressing these issues. The book also highlights the importance of large-scale agile development and provides guidance to organizations on how to transform their structure towards agile prioritization. The book covers various appropriate models, methodologies, and agile tools and

provides recommendations for cross-functional prioritization of requirements. It also considers the need for IT security and shows how it can be integrated into the overall agile development process.
Veracity of Data Laure Berti-Équille 2015-12-01 In the Web, a massive amount of user-generated contents are available through various channels (e.g., texts, tweets, Web tables, databases, multimedia-sharing platforms, etc.). Conflicting information, rumors, erroneous and fake contents can be easily spread across multiple sources, making it hard to distinguish between what is true and what is not. This monograph gives an overview of fundamental issues and recent contributions for ascertaining the veracity of data in the era of Big Data. The text is organized into six chapters, focusing on structured data extracted from texts. Chapter One introduces the problem of ascertaining the veracity of data in a multi-source and evolving context. Issues related to information extraction are presented in chapter Two. It is followed by practical techniques for evaluating data source reputation and authoritativeness in Chapter Three, including a review of the main models and Bayesian approaches of trust management. Current truth discovery computation algorithms are presented in details in Chapter Four. The theoretical foundations and various approaches for modeling diffusion phenomenon of misinformation spreading in networked systems is studied in Chapter Five. Finally, truth discovery computation from extracted data in a dynamic context of misinformation propagation raises interesting challenges that are explored in Chapter Six. Supplementary material including source codes, datasets, and slides are offered online. This text is intended for a seminar course at the graduate level. It is also to serve as a useful resource for researchers and practitioners who are interested in the study of fact-checking, truth discovery or rumor spreading.
SQL for Data Science Antonio Badia 2020-12-11 This textbook explains SQL within the context of data science and introduces the different parts of SQL as they are needed for the tasks usually carried out during data analysis. Using the framework of the data life cycle, it focuses on the steps that are very often given the short shift in traditional textbooks, like data loading, cleaning and pre-processing. The book is organized as follows. Chapter 1 describes the data life cycle, i.e. the sequence of stages from data acquisition to archiving, that data goes through as it is prepared and then actually analyzed, together with the different activities that take place at each stage. Chapter 2 gets into databases proper, explaining how relational databases organize data. Non-traditional data, like XML and text, are also covered. Chapter 3 introduces SQL queries, but unlike traditional textbooks, queries and their parts are described around typical data analysis tasks like data exploration, cleaning and transformation. Chapter 4 introduces some basic techniques for data analysis and shows how SQL can be used for some simple analyses without too much complication. Chapter 5 introduces additional SQL constructs that are important in a variety of situations and thus completes the coverage of SQL queries. Lastly, chapter 6 briefly explains how to use SQL from within R and from within Python programs. It focuses on how these languages can interact with a database, and how what has been learned about SQL can be leveraged to make life easier when using R or Python. All chapters contain a lot of examples and exercises on the way, and readers are encouraged to install the two open-source database systems (MySQL and Postgres) that are used throughout the book in order to practice and work on the exercises, because simply reading the book is much less useful than actually using it. This book is for anyone interested in data science and/or databases. It just demands a bit of computer fluency, but no specific background on databases or data analysis. All concepts are introduced intuitively and with a minimum of specialized jargon. After going through this book, readers should be able to profitably learn more about data mining, machine learning, and database management from more advanced textbooks and courses.
Semantics Empowered Web 3.0 Amit Sheth 2022-05-31 After the traditional document-centric Web 1.0 and user-generated content focused Web 2.0, Web 3.0 has become a repository of an ever growing variety of Web resources that include data and services associated with enterprises, social networks, sensors, cloud, as well as mobile and other devices that constitute the Internet of Things. These pose unprecedented challenges in terms of heterogeneity (variety), scale (volume), and continuous changes (velocity), as well as present corresponding opportunities if they can be exploited. Just as semantics has played a critical role in dealing with data heterogeneity in the past to provide interoperability and integration, it is playing an even more critical role in dealing with the challenges and helping users and applications exploit all forms of Web 3.0 data. This book presents a unified approach to harness and exploit all forms of contemporary Web resources using the core principles of ability to associate meaning with data through conceptual or domain models and semantic descriptions including annotations, and through advanced semantic techniques for search, integration, and analysis. It discusses the use of Semantic Web standards and techniques when appropriate, but also advocates the use of lighter weight, easier to use, and more scalable options when they are more suitable. The authors' extensive experience spanning research and prototypes to development of operational applications and commercial technologies and products guide the treatment of the material. Table of Contents: Role of Semantics and Metadata / Types and Models of Semantics / Annotation -- Adding Semantics to Data / Semantics for Enterprise Data / Semantics for Services / Semantics for Sensor Data / Semantics for Social Data / Semantics for Cloud Computing / Semantics for Advanced Applications
Database Systems for Advanced Applications Shamkant B. Navathe 2016-03-24 This two volume set LNCS 9642 and LNCS 9643 constitutes the refereed proceedings of the 21st International Conference on Database Systems for Advanced Applications, DASFAA 2016, held in Dallas, TX, USA, in April 2016. The 61 full papers presented were carefully reviewed and selected from a total of 183 submissions. The papers cover the following topics: crowdsourcing, data quality, entity identification, data mining and machine learning, recommendation, semantics computing and knowledge base, textual data, social networks, complex queries, similarity computing, graph databases, and miscellaneous, advanced applications.
Query Processing over Uncertain Databases Lei Chen 2022-05-31 Due to measurement errors, transmission lost, or injected noise for privacy protection, uncertainty exists in the data of many real applications. However, query processing techniques for deterministic data cannot be directly applied to uncertain data because they do not have mechanisms to handle data uncertainty. Therefore, efficient and effective manipulation of uncertain data is a practical yet challenging research topic. In this book, we start from the data models for imprecise and uncertain data, move on to defining different semantics for queries on uncertain data, and finally discuss the advanced query processing techniques for various probabilistic queries in uncertain databases. The book serves as a comprehensive guideline for query processing over uncertain databases. Table of Contents: Introduction / Uncertain Data Models / Spatial Query Semantics over Uncertain Data Models / Spatial Query Processing over Uncertain Databases / Conclusion
Frontiers in Statistical Quality Control 11 Sven Knoth 2015-04-24 The main focus of this edited volume is on three major areas of statistical quality control: statistical process control (SPC), acceptance sampling and design of experiments. The majority of the papers deal with statistical process control, while acceptance sampling and design of experiments are also treated to a lesser extent. The book is organized into four thematic parts, with Part I addressing statistical process control. Part II is devoted to acceptance sampling. Part III covers the design of experiments, while Part IV discusses related fields. The twenty-three papers in this volume stem from The 11th International Workshop on Intelligent Statistical Quality Control, which was held in Sydney, Australia from August 20 to August 23, 2013. The event was hosted by Professor Ross Sparks, CSIRO Mathematics, Informatics and Statistics, North Ryde, Australia and was jointly organized by Professors S. Knoth, W. Schmid and Ross Sparks. The papers presented here were carefully selected and reviewed by the scientific program committee, before being revised and adapted for this volume.
Handbook of Data Quality Shazia Sadiq 2013-08-13 The issue of data quality is as old as data itself. However, the proliferation of diverse, large-scale and often publically available data on the Web has increased the risk of poor data quality and misleading data interpretations. On the other hand, data is now exposed at a much more strategic level e.g. through business intelligence systems, increasing manifold the stakes involved for individuals, corporations as well as government agencies. There, the lack of knowledge about data accuracy, currency or completeness can have erroneous and even catastrophic results. With these changes, traditional approaches to data management in general, and data quality control specifically, are challenged. There is an evident need to incorporate data quality considerations into the whole data cycle, encompassing managerial/governance as well as technical aspects. Data quality experts from research and industry agree that a unified framework for data quality management should bring together organizational, architectural and computational approaches. Accordingly, Sadiq structured this handbook in four parts: Part I is on organizational solutions, i.e. the development of data quality objectives for the

organization, and the development of strategies to establish roles, processes, policies, and standards required to manage and ensure data quality. Part II, on architectural solutions, covers the technology landscape required to deploy developed data quality management processes, standards and policies. Part III, on computational solutions, presents effective and efficient tools and techniques related to record linkage, lineage and provenance, data uncertainty, and advanced integrity constraints. Finally, Part IV is devoted to case studies of successful data quality initiatives that highlight the various aspects of data quality in action. The individual chapters present both an overview of the respective topic in terms of historical research and/or practice and state of the art, as well as specific techniques, methodologies and frameworks developed by the individual contributors. Researchers and students of computer science, information systems, or business management as well as data professionals and practitioners will benefit most from this handbook by not only focusing on the various sections relevant to their research area or particular practical work, but by also studying chapters that they may initially consider not to be directly relevant to them, as there they will learn about new perspectives and approaches.

Conceptual Modeling Isabelle Comyn-Wattiau 2016-11-07 This book constitutes the refereed proceedings of the 345h International Conference on Conceptual Modeling, ER 2016, held in Gifu, Japan, in November 2016. The 23 full and 18 short papers presented together with 3 keynotes were carefully reviewed and selected from 113 submissions. The papers are organized in topical sections on Analytics and Conceptual Modeling; Conceptual Modeling and Ontologies; Requirements Engineering; Advanced Conceptual Modeling; Semantic Annotations; Modeling and Executing Business Processes; Business Process Management and Modeling; Applications and Experiments of Conceptual Modeling; Schema Mapping; Conceptual Modeling Guidance; and Goal Modeling.

Advances in Databases and Information Systems Alberto Abelló

Database Repairs and Consistent Query Answering Leopoldo Bertossi 2022-05-31 Integrity constraints are semantic conditions that a database should satisfy in order to be an appropriate model of external reality. In practice, and for many reasons, a database may not satisfy those integrity constraints, and for that reason it is said to be inconsistent. However, and most likely, a large portion of the database is still semantically correct, in a sense that has to be made precise. After having provided a formal characterization of consistent data in an inconsistent database, the natural problem emerges of extracting that semantically correct data, as query answers. The consistent data in an inconsistent database is usually characterized as the data that persists across all the database instances that are consistent and minimally differ from the inconsistent instance. Those are the so-called repairs of the database. In particular, the consistent answers to a query posed to the inconsistent database are those answers that can be simultaneously obtained from all the database repairs. As expected, the notion of repair requires an adequate notion of distance that allows for the comparison of databases with respect to how much they differ from the inconsistent instance. On this basis, the minimality condition on repairs can be properly formulated. In this monograph we present and discuss these fundamental concepts, different repair semantics, algorithms for computing consistent answers to queries, and also complexity-theoretic results related to the computation of repairs and doing consistent query answering. Table of Contents: Introduction / The Notions of Repair and Consistent Answer / Tractable CQA and Query Rewriting / Logically Specifying Repairs / Decision Problems in CQA: Complexity and Algorithms / Repairs and Data Cleaning

In Search of Elegance in the Theory and Practice of Computation Val Tannen 2013-10-28 This Festschrift volume, published in honour of Peter Buneman, contains contributions written by some of his colleagues, former students, and friends. In celebration of his distinguished career a colloquium was held in Edinburgh, Scotland, 27-29 October, 2013. The articles presented herein belong to some of the many areas of Peter's research interests.

Information Quality and Governance for Business Intelligence Yeoh, William 2013-12-31 Business intelligence initiatives have been dominating the technology priority list of many organizations. However, the lack of effective information quality and governance strategies and policies has been meeting these initiatives with some challenges. Information Quality and Governance for Business Intelligence presents the latest exchange of academic research on all aspects of practicing and managing information using a multidisciplinary approach that examines its quality for organizational growth. This book is an essential reference tool for researchers, practitioners, and university students specializing in business intelligence, information quality, and information systems.

Transactional Memory. Foundations, Algorithms, Tools, and Applications Rachid Guerraoui 2014-12-29 The advent of multi-core architectures and cloud-computing has brought parallel programming into the mainstream of software development. Unfortunately, writing scalable parallel programs using traditional lock-based synchronization primitives is well known to be a hard, time consuming and error-prone task, mastered by only a minority of specialized programmers. Building on the familiar abstraction of atomic transactions, Transactional Memory (TM) promises to free programmers from the complexity of conventional synchronization schemes, simplifying the development and verification of concurrent programs, enhancing code reliability, and boosting productivity. Over the last decade TM has been subject to intense research on a broad range of aspects including hardware and operating systems support, language integration, as well as algorithms and theoretical foundations. On the industrial side, the major players of the software and hardware markets have been up-front in the research and development of prototypal products providing support for TM systems. This has recently led to the introduction of hardware TM implementations on mainstream commercial microprocessors and to the integration of TM support for the world’s leading open source compiler. In such a vast inter-disciplinary domain, the Euro-TM COST Action (IC1001) has served as a catalyzer and a bridge for the various research communities looking at disparate, yet subtly interconnected, aspects of TM. This book emerged from the idea having Euro-TM experts compile recent results in the TM area in a single and consistent volume. Contributions have been carefully selected and revised to provide a broad coverage of several fundamental issues associated with the design and implementation of TM systems, including their theoretical underpinnings and algorithmic foundations, programming language integration and verification tools, hardware supports, distributed TM systems, self-tuning mechanisms, as well as lessons learnt from building complex TM-based applications.

Enterprise Big Data Engineering, Analytics, and Management Atzmueller, Martin 2016-06-01 The significance of big data can be observed in any decision-making process as it is often used for forecasting and predictive analytics. Additionally, big data can be used to build a holistic view of an enterprise through a collection and analysis of large data sets retrospectively. As the data deluge deepens, new methods for analyzing, comprehending, and making use of big data become necessary. Enterprise Big Data Engineering, Analytics, and Management presents novel methodologies and practical approaches to engineering, managing, and analyzing large-scale data sets with a focus on enterprise applications and implementation. Featuring essential big data concepts including data mining, artificial intelligence, and information extraction, this publication provides a platform for retargeting the current research available in the field. Data analysts, IT professionals, researchers, and graduate-level students will find the timely research presented in this publication essential to furthering their knowledge in the field.

New Horizons for a Data-Driven Economy José María Cavanillas 2016-04-04 In this book readers will find technological discussions on the existing and emerging technologies across the different stages of the big data value chain. They will learn about legal aspects of big data, the social impact, and about education needs and requirements. And they will discover the business perspective and how big data technology can be exploited to deliver value within different sectors of the economy. The book is structured in four parts: Part I “The Big Data Opportunity” explores the value potential of big data with a particular focus on the European context. It also describes the legal, business and social dimensions that need to be addressed, and briefly introduces the European Commission’s BIG project. Part II “The Big Data Value Chain” details the complete big data lifecycle from a technical point of view, ranging from data acquisition, analysis, curation and storage, to data usage and exploitation. Next, Part III “Usage and Exploitation of Big Data” illustrates the value creation possibilities of big data applications in various sectors, including industry, healthcare, finance, energy, media and public services. Finally, Part IV “A Roadmap for Big Data Research” identifies and prioritizes the cross-sectorial requirements for big data research, and outlines the most urgent and challenging technological, economic, political and societal issues for big data in Europe. This compendium summarizes more than two years of work performed by a leading group of major European research centers and industries in the context of the BIG project. It brings together research findings, forecasts and estimates related to this challenging technological context that is becoming the major axis of

the new digitally transformed business environment.
Entity Resolution in the Web of Data Vassilis Christophides 2015-08-01 In recent years, several knowledge bases have been built to enable large-scale knowledge sharing, but also an entity-centric Web search, mixing both structured data and text querying. These knowledge bases offer machine-readable descriptions of real-world entities, e.g., persons, places, published on the Web as Linked Data. However, due to the different information extraction tools and curation policies employed by knowledge bases, multiple, complementary and sometimes conflicting descriptions of the same real-world entities may be provided. Entity resolution aims to identify different descriptions that refer to the same entity appearing either within or across knowledge bases. The objective of this book is to present the new entity resolution challenges stemming from the openness of the Web of data in describing entities by an unbounded number of knowledge bases, the semantic and structural diversity of the descriptions provided across domains even for the same real-world entities, as well as the autonomy of knowledge bases in terms of adopted processes for creating and curating entity descriptions. The scale, diversity, and graph structuring of entity descriptions in the Web of data essentially challenge how two descriptions can be effectively compared for similarity, but also how resolution algorithms can efficiently avoid examining pairwise all descriptions. The book covers a wide spectrum of entity resolution issues at the Web scale, including basic concepts and data structures, main resolution tasks and workflows, as well as state-of-the-art algorithmic techniques and experimental trade-offs.
Data Exploration Using Example-Based Methods Matteo Lissandrini 2018-11-27 Data usually comes in a plethora of formats and dimensions, rendering the information extraction and exploration processes challenging. Thus, being able to perform exploratory analyses of the data with the intent of having an immediate glimpse of some of the data properties is becoming crucial. Exploratory analyses should be simple enough to avoid complicated declarative languages (such as SQL) and mechanisms, while at the same time retaining the flexibility and expressiveness of such languages. Recently, we have witnessed a rediscovery of the so-called example-based methods, in which the user, or analyst, circumvents query languages by using examples as input. An example is a representative of the intended results or, in other words, an item from the result set. Example-based methods exploit inherent characteristics of the data to infer the results that the user has in mind but may not be able to (easily) express. They can be useful in cases where a user is looking for information in an unfamiliar dataset, when they are performing a particularly challenging task like finding duplicate items, or when they are simply exploring the data. In this book, we present an excursus over the main methods for exploratory analysis, with a particular focus on example-based methods. We show how different data types require different techniques and present algorithms that are specifically designed for relational, textual, and graph data. The book also presents the challenges and new frontiers of machine learning in online settings that have recently attracted the attention of the database community. The book concludes with a vision for further research and applications in this area.
Foundations of Data Quality Management Wenfei Fan 2022-05-31 Data quality is one of the most important problems in data management. A database system typically aims to support the creation, maintenance, and use of large amount of data, focusing on the quantity of data. However, real-life data are often dirty: inconsistent, duplicated, inaccurate, incomplete, or stale. Dirty data in a database routinely generate misleading or biased analytical results and decisions, and lead to loss of revenues, credibility and customers. With this comes the need for data quality management. In contrast to traditional data management tasks, data quality management enables the detection and correction of errors in the data, syntactic or semantic, in order to improve the quality of the data and hence, add value to business processes. While data quality has been a longstanding problem for decades, the prevalent use of the Web has increased the risks, on an unprecedented scale, of creating and propagating dirty data. This monograph gives an overview of fundamental issues underlying central aspects of data quality, namely, data consistency, data deduplication, data accuracy, data currency, and information completeness. We promote a uniform logical framework for dealing with these issues, based on data quality rules. The text is organized into seven chapters, focusing on relational data. Chapter One introduces data quality issues. A conditional dependency theory is developed in Chapter Two, for capturing data inconsistencies. It is followed by practical techniques in Chapter 2b for discovering conditional dependencies, and for detecting inconsistencies and repairing data based on conditional dependencies. Matching dependencies are introduced in Chapter Three, as matching rules for data deduplication. A theory of relative information completeness is studied in Chapter Four, revising the classical Closed World Assumption and the Open World Assumption, to characterize incomplete information in the real world. A data currency model is presented in Chapter Five, to identify the current values of entities in a database and to answer queries with the current values, in the absence of reliable timestamps. Finally, interactions between these data quality issues are explored in Chapter Six. Important theoretical results and practical algorithms are covered, but formal proofs are omitted. The bibliographical notes contain pointers to papers in which the results were presented and proven, as well as references to materials for further reading. This text is intended for a seminar course at the graduate level. It is also to serve as a useful resource for researchers and practitioners who are interested in the study of data quality. The fundamental research on data quality draws on several areas, including mathematical logic, computational complexity and database theory. It has raised as many questions as it has answered, and is a rich source of questions and vitality. Table of Contents: Data Quality: An Overview / Conditional Dependencies / Cleaning Data with Conditional Dependencies / Data Deduplication / Information Completeness / Data Currency / Interactions between Data Quality Issues
Handbook of Research on Machine Learning Innovations and Trends Hassanien, Aboul Ella 2017-04-03 Continuous improvements in technological applications have allowed more opportunities to develop automated systems. This not only leads to higher success in smart data analysis, but it increases the overall probability of technological progression. The Handbook of Research on Machine Learning Innovations and Trends is a key resource on the latest advances and research regarding the vast range of advanced systems and applications involved in machine intelligence. Highlighting multidisciplinary studies on decision theory, intelligent search, and multi-agent systems, this publication is an ideal reference source for professionals and researchers working in the field of machine learning and its applications.
Data Cleaning Venkatesh Ganti 2022-05-31 Data warehouses consolidate various activities of a business and often form the backbone for generating reports that support important business decisions. Errors in data tend to creep in for a variety of reasons. Some of these reasons include errors during input data collection and errors while merging data collected independently across different databases. These errors in data warehouses often result in erroneous upstream reports, and could impact business decisions negatively. Therefore, one of the critical challenges while maintaining large data warehouses is that of ensuring the quality of data in the data warehouse remains high. The process of maintaining high data quality is commonly referred to as data cleaning. In this book, we first discuss the goals of data cleaning. Often, the goals of data cleaning are not well defined and could mean different solutions in different scenarios. Toward clarifying these goals, we abstract out a common set of data cleaning tasks that often need to be addressed. This abstraction allows us to develop solutions for these common data cleaning tasks. We then discuss a few popular approaches for developing such solutions. In particular, we focus on an operator-centric approach for developing a data cleaning platform. The operator-centric approach involves the development of customizable operators that could be used as building blocks for developing common solutions. This is similar to the approach of relational algebra for query processing. The basic set of operators can be put together to build complex queries. Finally, we discuss the development of custom scripts which leverage the basic data cleaning operators along with relational operators to implement effective solutions for data cleaning tasks.
The Fundamentals of Quality Management D.F. Kehoe 2012-12-06 This book has been written to provide both students and industrial man agers with a comprehensive description of the tools and techniques of Quality Management and also to provide a framework for understanding Quality Development. Central to the theme of this book is the idea that quality management is a developmental process which requires an understanding of the techniques, the people and the systems issues. The aims of quality development are to produce greater organizational consistency, to improve customer satisfac tion and to reduce the business process costs. In order to achieve these aims, managers are required to have an understanding of both the underlying the ories and the methodologies for implementation. The aim of this book is to provide a

coherent description of both the theoretical and implementation aspects of quality management. Since the halcyon days of the quality 'revolution' of the 1970s and 1980s, many organizations have realized that quality development represents an enormous management challenge. This challenge for continuous improve ment requires the continuous development of systems, of techniques and of people. Like most serious business strategies, competitive improvement through quality development can only be achieved if the organization understands not only what the various quality 'options' are but also when a particular technique or approach is applicable. Quality development has no single blueprint but requires a learning organization which understands key concepts and methods of implementation.

Linked Data Management Andreas Harth 2016-04-19 Linked Data Management presents techniques for querying and managing Linked Data that is available on today’s Web. The book shows how the abundance of Linked Data can serve as fertile ground for research and commercial applications. The text focuses on aspects of managing large-scale collections of Linked Data. It offers a detailed introduction to Linked Data and related standards, including the main principles distinguishing Linked Data from standard database technology. Chapters also describe how to generate links between datasets and explain the overall architecture of data integration systems based on Linked Data. A large part of the text is devoted to query processing in different setups. After presenting methods to publish relational data as Linked Data and efficient centralized processing, the book explores lookup-based, distributed, and parallel solutions. It then addresses advanced topics, such as reasoning, and discusses work related to read-write Linked Data for system interoperation. Despite the publication of many papers since Tim Berners-Lee developed the Linked Data principles in 2006, the field lacks a comprehensive, unified overview of the state of the art. Suitable for both researchers and practitioners, this book provides a thorough, consolidated account of the new data publishing and data integration paradigm. While the book covers query processing extensively, the Linked Data abstraction furnishes more than a mechanism for collecting, integrating, and querying data from the open Web—the Linked Data technology stack also allows for controlled, sophisticated applications deployed in an enterprise environment.

Information Quality in Information Fusion and Decision Making Éloi Bossé 2019-04-02 This book presents a contemporary view of the role of information quality in information fusion and decision making, and provides a formal foundation and the implementation strategies required for dealing with insufficient information quality in building fusion systems for decision making. Information fusion is the process of gathering, processing, and combining large amounts of information from multiple and diverse sources, including physical sensors to human intelligence reports and social media. That data and information may be unreliable, of low fidelity, insufficient resolution, contradictory, fake and/or redundant. Sources may provide unverified reports obtained from other sources resulting in correlations and biases. The success of the fusion processing depends on how well knowledge produced by the processing chain represents reality, which in turn depends on how adequate data are, how good and adequate are the models used, and how accurate, appropriate or applicable prior and contextual knowledge is. By offering contributions by leading experts, this book provides an unparalleled understanding of the problem of information quality in information fusion and decision-making for researchers and professionals in the field.

The Algorithmic Foundations of Differential Privacy Cynthia Dwork 2014 The problem of privacy-preserving data analysis has a long history spanning multiple disciplines. As electronic data about individuals becomes increasingly detailed, and as technology enables ever more powerful collection and curation of these data, the need increases for a robust, meaningful, and mathematically rigorous definition of privacy, together with a computationally rich class of algorithms that satisfy this definition. Differential Privacy is such a definition. The Algorithmic Foundations of Differential Privacy starts out by motivating and discussing the meaning of differential privacy, and proceeds to explore the fundamental techniques for achieving differential privacy, and the application of these techniques in creative combinations, using the query-release problem as an ongoing example. A key point is that, by rethinking the computational goal, one can often obtain far better results than would be achieved by methodically replacing each step of a non-private computation with a differentially private implementation. Despite some powerful computational results, there are still fundamental limitations. Virtually all the algorithms discussed herein maintain differential privacy against adversaries of arbitrary computational power -- certain algorithms are computationally intensive, others are efficient. Computational complexity for the adversary and the algorithm are both discussed. The monograph then turns from fundamentals to applications other than query-release, discussing differentially private methods for mechanism design and machine learning. The vast majority of the literature on differentially private algorithms considers a single, static, database that is subject to many analyses. Differential privacy in other models, including distributed databases and computations on data streams, is discussed. The Algorithmic Foundations of Differential Privacy is meant as a thorough introduction to the problems and techniques of differential privacy, and is an invaluable reference for anyone with an interest in the topic.

Foundations of Data Quality Management Wenfei Fan 2012 Data quality is one of the most important problems in data management. A database system typically aims to support the creation, maintenance and use of large amount of data, focusing on the quantity of data. However, real-life data are often dirty: inconsistent, duplicated, inaccurate, incomplete, or stale. Dirty data in a database routinely generate misleading or biased analytical results and decisions, and lead to loss of revenues, credibility and customers. With this comes the need for data quality management. In contrast to traditional data management tasks, data quality management is to enable the detection and correction of errors in the data, syntactic or semantic, in order to improve the quality of the data and hence, add values to business processes. This monograph gives an overview of fundamental issues underlying central aspects of data quality, namely, data consistency, deduplication, accuracy, currency, and information completeness. We promote a uniform logical framework for dealing with these issues, based on data quality rules. The text is organized into seven chapters, focusing on relational data. Chapter 1 introduces data quality issues. A conditional dependency theory is developed in Chapter 2, for capturing data inconsistencies. It is followed by practical techniques in Chapter 3 for discovering conditional dependencies, and for detecting inconsistencies and repairing data based on conditional dependencies. Matching dependencies are introduced in Chapter 4, as matching rules for data deduplication. A theory of relative information completeness is studied in Chapter 5, revising the classical Closed World Assumption and the Open World Assumption, to characterize incomplete information in the real world. A data currency model is presented in Chapter 6, to identify the current values of entities in a database and to answer queries with the current values, in the absence of reliable timestamps. Finally, interactions between these data quality issues are explored in Chapter 7. Important theoretical results and practical algorithms are covered, but formal proofs are omitted. The bibliographical notes contain pointers to papers in which the results were presented and proved, as well as references to materials for further reading. This text is intended for a seminar course at the graduate level. It is also to serve as a useful resource for researchers and practitioners who are interested in the study of data quality. The fundamental research on data quality draws on several areas, including mathematical logic, computational complexity and database theory. It has raised as many questions as it has answered, and is a rich source of questions and vitality.

Advances in System-Integrated Intelligence Maurizio Valle 2022-09-03 This book reports on cutting-edge research and developments focusing on integrating intelligent functionalities into materials, components, systems and products. Gathering the proceedings of the 6th International Conference on System-Integrated Intelligence (SysInt 2022), held on September 7-9, in Genova, Italy, it offers a comprehensive, multidisciplinary and applied perspective on the state-of-the art and challenges in the field of intelligent, flexible and connected systems. The book covers advanced methods and applications relating to artificial, pervasive and ubiquitous intelligence, sensors, smart factory and logistics, structural health monitoring, as well as soft robotics, cognitive systems and human-machine interaction. Giving a special focus to artificial intelligence, it extensively reports on methods and algorithms for data-driven modeling, and agent-based data processing and planning. It aims at inspiring and fostering collaboration between researchers and professionals from the different fields of electrical, manufacturing and production engineering, and materials and computer sciences.