

Apache Spark Book Soup

Reveals how the voyages of Columbus reintroduced plants and animals that had been separated millions of years earlier, documenting how the ensuing exchange of flora and fauna between Eurasia and the Americas fostered a European rise, decimated imperial China and rendered Manila and Mexico City the center of the world for two centuries. Reprint.

Completion of the first phase of the Human Genome Project has presented scientists with a mountain of new information. The availability of all human genes and their locations is exciting, but their mechanisms of action and interaction with other genes are often unknown. Certain variations in the environment and characteristics of human genes make it difficult for scientists to maximize their efforts to take advantage of this new fund of knowledge. The high degree of similarity between mouse and human genomes and physiology has made mice the laboratory animals of choice in the search for models of human disease. In recent years significant advances in our ability to examine the mouse eye and to view vascular and neural structures in vivo has led to a marked increase in interest in mouse ocular pathology and development. Besides providing much new information, *Systematic Evaluation of the Mouse Eye: Anatomy, Pathology, and Biomethods* provides investigators with a resource that summarizes the widely scattered information on mouse eye anatomy, development, and pathology. It describes the normal anatomy, pre- and postnatal development, regional pathology, and methodology for evaluation of the mouse eye and adnexae. The first section covers the principal features of mouse eye anatomy and development. The second part discusses issues in evaluation of mutations in mice, including background ocular disease, control selection, quantitative traits,

Read Book Apache Spark Book Soup

and ENU mutagenesis. With this background in place, the book then reviews and illustrates regional ocular pathology and correlates it with human eye disease. Many of the contributors who have helped develop and refine the specialized techniques necessary for the examination and preparation of mouse eye tissue for microscopy provide "how-to" chapters. These chapters cover methods for necropsy, photography, histopathology, confocal microscopy, intraocular pressure measurement, and electroretinography. *Systematic Evaluation of the Mouse Eye: Anatomy, Pathology, and Biomethods* provides the first comprehensive and systematic guide to understanding the development, structure, and abnormal pathology of the mouse eye so you can make the most of your genetic research.

Learn web scraping and crawling techniques to access unlimited data from any web source in any format. With this practical guide, you'll learn how to use Python scripts and web APIs to gather and process data from thousands—or even millions—of web pages at once. Ideal for programmers, security professionals, and web administrators familiar with Python, this book not only teaches basic web scraping mechanics, but also delves into more advanced topics, such as analyzing raw data or using scrapers for frontend website testing. Code samples are available to help you understand the concepts in practice. Learn how to parse complicated HTML pages Traverse multiple pages and sites Get a general overview of APIs and how they work Learn several methods for storing the data you scrape Download, read, and extract data from documents Use tools and techniques to clean badly formatted data Read and write natural languages Crawl through forms and logins Understand how to scrape JavaScript Learn image processing and text recognition

From Chile-Glazed Pork Chops to Dos Leches Flan, Lourdes Castro offers authentic, no-fuss

Read Book Apache Spark Book Soup

Mexican meals with clean, vibrant flavors that are the essence of great Mexican food. In *Simply Mexican*, Castro presents authentic recipes that don't require a fortnight to prepare or extended shopping forays to find rare ingredients. Castro honed her knowledge of traditional south-of-the-border dishes by teaching the fundamentals to adults and children at her Miami cooking school, and now she's introducing real Mexican fare that works for busy cooks every night of the week. *Simply Mexican* features easy-to-prepare, fun-to-eat favorites with big flavors, such as Chicken Enchiladas with Tomatillo Sauce and Crab Tostadas. Once you have mastered the basics, Castro will guide you through more advanced Mexican mainstays such as adobo and mole, and show you how to make the most effortless savory and sweet tamales around. With cooking notes that highlight useful equipment, new ingredients, shortcut techniques, and instructions for advance preparation, *Simply Mexican* demystifies authentic Mexican meals so you can make them at home in a snap. "With this book Lourdes Castro has added a spark of creativity and simplicity to Mexican food that up until now had not yet been realized. It gives me a huge sense of pride and honor to know that this book exists, as it will help a large audience re-create these gems in a simple and straightforward way." —Aarón Sánchez, chef/owner of Paladar and chef/partner of Centrico, author of *La Comida del Barrio*, and former cohost of Food Network's *Melting Pot* "Here, at last, are real Mexican recipes that are authentic, creative, and fun to prepare. Lourdes Castro creates an atmosphere that makes learning about enchiladas, tacos, and salsas exciting and interesting, and her precise methodology with Mexican cookery is refreshing and very entertaining. Highly recommended." —Jonathan Waxman, chef/owner of Barbuto and West County Grill and author of *A Great American Cook*

Read Book Apache Spark Book Soup

Uses picture books to teach such key scientific topics as seeds and plants, habitats, light and color, and water cycle and includes science information, management tips, book connections, Web sites, and assessment ideas.

Dr. Greg Zacharias, former Chief Scientist of the United States Air Force (2015-18), explores next steps in autonomous systems (AS) development, fielding, and training. Rapid advances in AS development and artificial intelligence (AI) research will change how we think about machines, whether they are individual vehicle platforms or networked enterprises. The payoff will be considerable, affording the US military significant protection for aviators, greater effectiveness in employment, and unlimited opportunities for novel and disruptive concepts of operations. *Autonomous Horizons: The Way Forward* identifies issues and makes recommendations for the Air Force to take full advantage of this transformational technology. *Summary Spring Batch in Action* is an in-depth guide to writing batch applications using Spring Batch. Written for developers who have basic knowledge of Java and the Spring lightweight container, the book provides both a best-practices approach to writing batch jobs and comprehensive coverage of the Spring Batch framework. *About the Technology* Even though running batch jobs is a common task, there's no standard way to write them. Spring Batch is a framework for writing batch applications in Java. It includes reusable components and a solid runtime environment, so you don't have to start a new project from scratch. And it uses Spring's familiar programming model to simplify configuration and implementation, so it'll be comfortably familiar to most Java developers. *About the Book* *Spring Batch in Action* is a thorough, in-depth guide to writing efficient batch applications. Starting with the basics, it discusses the best practices of batch jobs along with details of the Spring Batch framework.

Read Book Apache Spark Book Soup

You'll learn by working through dozens of practical, reusable examples in key areas like monitoring, tuning, enterprise integration, and automated testing. No prior batch programming experience is required. Basic knowledge of Java and Spring is assumed. Purchase of the print book comes with an offer of a free PDF, ePub, and Kindle eBook from Manning. Also available is all code from the book. What's Inside Batch programming from the ground up Implementing data components Handling errors during batch processing Automating tedious tasks Table of Contents PART 1 BACKGROUND Introducing Spring Batch Spring Batch concepts PART 2 CORE SPRING BATCH Batch configuration Running batch jobs Reading data Writing data Processing data Implementing bulletproof jobs Transaction management PART 3 ADVANCED SPRING BATCH Controlling execution Enterprise integration Monitoring jobs Scaling and parallel processing Testing batch applications

Get Started Fast with Apache Hadoop® 2, YARN, and Today's Hadoop Ecosystem With Hadoop 2.x and YARN, Hadoop moves beyond MapReduce to become practical for virtually any type of data processing. Hadoop 2.x and the Data Lake concept represent a radical shift away from conventional approaches to data usage and storage. Hadoop 2.x installations offer unmatched scalability and breakthrough extensibility that supports new and existing Big Data analytics processing methods and models. Hadoop® 2 Quick-Start Guide is the first easy, accessible guide to Apache Hadoop 2.x, YARN, and the modern Hadoop ecosystem. Building on his unsurpassed experience teaching Hadoop and Big Data, author Douglas Eadline covers all the basics you need to know to install and use Hadoop 2 on personal computers or servers, and to navigate the powerful technologies that complement it. Eadline concisely introduces and explains every key Hadoop 2 concept, tool, and service, illustrating each with a simple

Read Book Apache Spark Book Soup

“beginning-to-end” example and identifying trustworthy, up-to-date resources for learning more. This guide is ideal if you want to learn about Hadoop 2 without getting mired in technical details. Douglas Eadline will bring you up to speed quickly, whether you’re a user, admin, devops specialist, programmer, architect, analyst, or data scientist. Coverage Includes Understanding what Hadoop 2 and YARN do, and how they improve on Hadoop 1 with MapReduce Understanding Hadoop-based Data Lakes versus RDBMS Data Warehouses Installing Hadoop 2 and core services on Linux machines, virtualized sandboxes, or clusters Exploring the Hadoop Distributed File System (HDFS) Understanding the essentials of MapReduce and YARN application programming Simplifying programming and data movement with Apache Pig, Hive, Sqoop, Flume, Oozie, and HBase Observing application progress, controlling jobs, and managing workflows Managing Hadoop efficiently with Apache Ambari—including recipes for HDFS to NFSv3 gateway, HDFS snapshots, and YARN configuration Learning basic Hadoop 2 troubleshooting, and installing Apache Hue and Apache Spark

Frank Kane's hands-on Spark training course, based on his bestselling Taming Big Data with Apache Spark and Python video, now available in a book. Understand and analyze large data sets using Spark on a single system or on a cluster. About This Book Understand how Spark can be distributed across computing clusters Develop and run Spark jobs efficiently using Python A hands-on tutorial by Frank Kane with over 15 real-world examples teaching you Big Data processing with Spark Who This Book Is For If you are a data scientist or data analyst who wants to learn Big Data processing

Read Book Apache Spark Book Soup

using Apache Spark and Python, this book is for you. If you have some programming experience in Python, and want to learn how to process large amounts of data using Apache Spark, Frank Kane's *Taming Big Data with Apache Spark and Python* will also help you. What You Will Learn Find out how you can identify Big Data problems as Spark problems Install and run Apache Spark on your computer or on a cluster Analyze large data sets across many CPUs using Spark's Resilient Distributed Datasets Implement machine learning on Spark using the MLlib library Process continuous streams of data in real time using the Spark streaming module Perform complex network analysis using Spark's GraphX library Use Amazon's Elastic MapReduce service to run your Spark jobs on a cluster In Detail Frank Kane's *Taming Big Data with Apache Spark and Python* is your companion to learning Apache Spark in a hands-on manner. Frank will start you off by teaching you how to set up Spark on a single system or on a cluster, and you'll soon move on to analyzing large data sets using Spark RDD, and developing and running effective Spark jobs quickly using Python. Apache Spark has emerged as the next big thing in the Big Data domain – quickly rising from an ascending technology to an established superstar in just a matter of years. Spark allows you to quickly extract actionable insights from large amounts of data, on a real-time basis, making it an essential tool in many modern businesses. Frank has packed this book with over 15 interactive, fun-filled examples relevant to the real world, and he will empower you to understand the Spark ecosystem and implement production-grade

Read Book Apache Spark Book Soup

real-time Spark projects with ease. Style and approach Frank Kane's *Taming Big Data with Apache Spark and Python* is a hands-on tutorial with over 15 real-world examples carefully explained by Frank in a step-by-step manner. The examples vary in complexity, and you can move through them at your own pace.

Summary *Introducing Data Science* teaches you how to accomplish the fundamental tasks that occupy data scientists. Using the Python language and common Python libraries, you'll experience firsthand the challenges of dealing with data at scale and gain a solid foundation in data science. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Many companies need developers with data science skills to work on projects ranging from social media marketing to machine learning. Discovering what you need to learn to begin a career as a data scientist can seem bewildering. This book is designed to help you get started. About the Book *Introducing Data*

Science *Introducing Data Science* explains vital data science concepts and teaches you how to accomplish the fundamental tasks that occupy data scientists. You'll explore data visualization, graph databases, the use of NoSQL, and the data science process. You'll use the Python language and common Python libraries as you experience firsthand the challenges of dealing with data at scale. Discover how Python allows you to gain insights from data sets so big that they need to be stored on multiple machines, or from data moving so quickly that no single machine can handle it. This book gives

Read Book Apache Spark Book Soup

you hands-on experience with the most popular Python data science libraries, Scikit-learn and StatsModels. After reading this book, you'll have the solid foundation you need to start a career in data science. What's Inside Handling large data Introduction to machine learning Using Python to work with data Writing data science algorithms About the Reader This book assumes you're comfortable reading code in Python or a similar language, such as C, Ruby, or JavaScript. No prior experience with data science is required. About the Authors Davy Cielen, Arno D. B. Meysman, and Mohamed Ali are the founders and managing partners of Optimately and Maiton, where they focus on developing data science projects and solutions in various sectors. Table of Contents Data science in a big data world The data science process Machine learning Handling large data on a single computer First steps in big data Join the NoSQL movement The rise of graph databases Text mining and text analytics Data visualization to the end user

Work with petabyte-scale datasets while building a collaborative, agile workplace in the process. This practical book is the canonical reference to Google BigQuery, the query engine that lets you conduct interactive analysis of large datasets. BigQuery enables enterprises to efficiently store, query, ingest, and learn from their data in a convenient framework. With this book, you'll examine how to analyze data at scale to derive insights from large datasets efficiently. Valliappa Lakshmanan, tech lead for Google Cloud Platform, and Jordan Tigani, engineering director for the BigQuery team, provide

Read Book Apache Spark Book Soup

best practices for modern data warehousing within an autoscaled, serverless public cloud. Whether you want to explore parts of BigQuery you're not familiar with or prefer to focus on specific tasks, this reference is indispensable.

“Atkinson and Jewell invite each of us to reimagine one’s connection to the land while cultivating nature close to home. A must-read for anyone searching for inspired solutions for designing or refining a garden.” —Emily Murphy, founder of Pass the Pistil

From windswept deserts to misty seaside hills and verdant valleys, the natural landscapes of the American West offer an astounding variety of climates for gardens. *Under Western Skies* reveals thirty-six of the most innovative designs—all embracing and celebrating the very soul of the land on which they grow. For the gardeners featured here, nature is the ultimate inspiration rather than something to be dominated, and *Under Western Skies* shows the strong connection each garden has with its place. Packed with Atkinson’s stunning photographs and illuminated by Jewell’s deep interest in the relationships between people and the spaces they inhabit, *Under Western Skies* offers page after page of encouraging ingenuity and inventive design for passionate gardeners who call the West home.

California Friendly® is California's future. Water reliability is dependent on using water wisely. We need to create sustainable gardens that rely on less water. This maintenance guide will help you support California's future.*Uncover the secrets of efficient irrigation.*Explore the techniques for irrigating with recycled water.*Get the

Read Book Apache Spark Book Soup

maintenance tips for hundreds of California Friendly® plants.*Discover the methods and means of managing weed and pest infestations.*Learn how to maintain rainwater capture opportunities.This book has been written for every landscaper, gardener and land manager in Southern California. It has been produced by the very first collaboration between three Southern California organizations, LADWP, MWD and SoCalGas. Grab a copy--they are free--use the information in your garden and help us create a beautiful, thriving and sustainable future.

The award-winning creator of Blue's Clues, Super Why!, and Daniel Tiger's Neighborhood shares the secret sauce behind her shows' powerful, transformative results in the form of eleven research-based, foundational "clues" to ensure that preschoolers flourish academically, socially, and emotionally during this critical time. The preschool years—when children are between the ages of two and five—are the most influential, important years in a child's life. Studies show that pausing to interact, playing to solve problems, diffusing with humor, and using repetition are the hidden clues conscious parents use to raise successful kids and help them learn critical thinking skills, foster empathy, and nurture their sense of self-worth. Angela C. Santomero, MA, the award-winning creator of children's television phenomena knows this better than anyone and has spent decades working to instill confidence in her young viewers. In *Preschool Clues*, she breaks down the philosophy behind her shows—educating, inspiring, and empowering kids—into concrete strategies that parents

Read Book Apache Spark Book Soup

and educators can incorporate into their family and classroom to set their preschoolers up for success, such as:

- Intentionally pausing to foster bonding, independence, and resilience
- Developing empathy and confidence through soliciting preschoolers' help
- Becoming "fluent" in the language of preschoolers: Play
- Igniting your preschooler's curiosity
- Being an involved co-player everyday
- Designing a healthy media diet

In *Preschool Clues*, Angela shares the latest research from top thinkers in child development and education. Through her practical, straightforward advice and inspiring, conversational approach, you will not only understand exactly what your children are learning from the shows they watch and why these shows are so effective, you'll know exactly how to apply these same proven approaches in your daily life and with the same powerful results.

Summary Natural Language Processing in Action is your guide to creating machines that understand human language using the power of Python with its ecosystem of packages dedicated to NLP and AI. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology

Recent advances in deep learning empower applications to understand text and speech with extreme accuracy. The result? Chatbots that can imitate real people, meaningful resume-to-job matches, superb predictive search, and automatically generated document summaries—all at a low cost. New techniques, along with accessible tools like Keras and TensorFlow, make professional-quality NLP easier than ever before. About

Read Book Apache Spark Book Soup

the Book Natural Language Processing in Action is your guide to building machines that can read and interpret human language. In it, you'll use readily available Python packages to capture the meaning in text and react accordingly. The book expands traditional NLP approaches to include neural networks, modern deep learning algorithms, and generative techniques as you tackle real-world problems like extracting dates and names, composing text, and answering free-form questions. What's inside

Some sentences in this book were written by NLP! Can you guess which ones?

Working with Keras, TensorFlow, gensim, and scikit-learn Rule-based and data-based NLP Scalable pipelines About the Reader This book requires a basic understanding of deep learning and intermediate Python skills. About the Author Hobson Lane, Cole Howard, and Hannes Max Hapke are experienced NLP engineers who use these techniques in production. Table of Contents

PART 1 - WORDY MACHINES Packets of thought (NLP overview) Build your vocabulary (word tokenization) Math with words (TF-IDF vectors) Finding meaning in word counts (semantic analysis)

PART 2 - DEEPER LEARNING (NEURAL NETWORKS) Baby steps with neural networks (perceptrons and backpropagation) Reasoning with word vectors (Word2vec) Getting words in order with convolutional neural networks (CNNs) Loopy (recurrent) neural networks (RNNs) Improving retention with long short-term memory networks Sequence-to-sequence models and attention

PART 3 - GETTING REAL (REAL-WORLD NLP CHALLENGES) Information extraction (named entity extraction and question answering) Getting chatty

Read Book Apache Spark Book Soup

(dialog engines) Scaling up (optimization, parallelization, and batch processing)

Journalist Walls grew up with parents whose ideals and stubborn nonconformity were their curse and their salvation. Rex and Rose Mary and their four children lived like nomads, moving among Southwest desert towns, camping in the mountains. Rex was a charismatic, brilliant man who, when sober, captured his children's imagination, teaching them how to embrace life fearlessly. Rose Mary painted and wrote and couldn't stand the responsibility of providing for her family. When the money ran out, the Walls retreated to the dismal West Virginia mining town Rex had tried to escape. As the dysfunction escalated, the children had to fend for themselves, supporting one another as they found the resources and will to leave home. Yet Walls describes her parents with deep affection in this tale of unconditional love in a family that, despite its profound flaws, gave her the fiery determination to carve out a successful life. -- From publisher description.

Do you have a real relationship with God, or do you just have a religion? Do you know God, or do you just know about God? In *How Big Is Your God?* Paul Coutinho, SJ, challenges us to grow stronger and deeper in our faith and in our relationship with God—a God whose love knows no bounds. To help us on our way, Coutinho introduces us to people in various world religions—from Hindu friends to Buddhist teachers to St. Ignatius of Loyola—who have shaped his spiritual life and made possible his deep, personal relationship with God.

Read Book Apache Spark Book Soup

This book is a practical, hands-on guide that takes you through the techniques of web scraping using Beautiful Soup. Getting Started with Beautiful Soup is great for anybody who is interested in website scraping and extracting information. However, a basic knowledge of Python, HTML tags, and CSS is required for better understanding.

Playing in puddles is an enjoyable experience that children love to do! Splashing in the water, building puddle boats, and using the water to make 'puddle soup' are all favourite activities that also integrate rich learning. Heading outside to play on a wet, rainy day uses children's excitement and enthusiasm for sensory experiences to inspire rich mathematical and scientific inquiries about what they are seeing and experiencing. In this book readers explore the math that exists in puddle play using the 'see, think, and wonder' thinking routine. Colourful photos and an easy to access text invite both children and adults to explore the fun and excitement of playing with water outdoors. This book can spark mathematical conversations with children, and be used as a guide for discovering the rich math that exists in weather-based activities. Photos can also be used without text to engage children in math talks as they observe and discuss what they see on each page. Educators will find this book to be an effective resource to inspire math talks within the classroom or before heading outside to play. Often educators are asked to justify messy outdoor play in the puddles. Exploring Puddle Math: See, Think and Wonder before heading outside will help empower children to notice and name the learning they are experiencing while having fun in the water.

Read Book Apache Spark Book Soup

Families can use this book as a tool for helping inspire math thinking and conversation with children during quality reading time together or before heading outside to play. One of the preeminent political philosophers of the 20th century makes a compelling argument for the political cogency of literary writing in this book which among to his intellectual autobiography and an introduction to his work.

Hungry? Got a craving for some ooey, gooey slime? Learn how to master the art of culinary slime making with YouTuber and slime Chef, Krist Soup. Included is a full menu of over 30 DIY slime meal recipes for Breakfast, Lunch, Dinner, and Dessert. Each recipe comes with step by step instructions and guided images to help relieve stress, enhance creativity, and bring some serious satisfaction to the brain and soul. MADE WITH LOVE AND ALL NATURAL GLUE. "Slime Soup" covers how to make basic slime and various types of some of the most popular slime. (Jiggly Slime, Butter Slime, Crunchy Slime, Fluffy Slime, etc.) "I created this book to spark creativity in all individuals. Not only do I want people to have fun while making slime, but also for people to get interested in learning how to cook meals. Plus, it's the perfect excuse to play with your food."

Learn the art of efficient web scraping and crawling with Python About This Book Extract data from any source to perform real time analytics. Full of techniques and examples to help you crawl websites and extract data within hours. A hands-on guide to web scraping and crawling with real-life problems and solutions Who This Book Is For If

Read Book Apache Spark Book Soup

you are a software developer, data scientist, NLP or machine-learning enthusiast or just need to migrate your company's wiki from a legacy platform, then this book is for you. It is perfect for someone , who needs instant access to large amounts of semi-structured data effortlessly. What You Will Learn Understand HTML pages and write XPath to extract the data you need Write Scrapy spiders with simple Python and do web crawls Push your data into any database, search engine or analytics system Configure your spider to download files, images and use proxies Create efficient pipelines that shape data in precisely the form you want Use Twisted Asynchronous API to process hundreds of items concurrently Make your crawler super-fast by learning how to tune Scrapy's performance Perform large scale distributed crawls with scrapyd and scrapinghub In Detail This book covers the long awaited Scrapy v 1.0 that empowers you to extract useful data from virtually any source with very little effort. It starts off by explaining the fundamentals of Scrapy framework, followed by a thorough description of how to extract data from any source, clean it up, shape it as per your requirement using Python and 3rd party APIs. Next you will be familiarised with the process of storing the scrapped data in databases as well as search engines and performing real time analytics on them with Spark Streaming. By the end of this book, you will perfect the art of scarping data for your applications with ease Style and approach It is a hands on guide, with first few chapters written as a tutorial, aiming to motivate you and get you started quickly. As the book progresses, more advanced features are explained with

Read Book Apache Spark Book Soup

real world examples that can be referred while developing your own web applications. This latest edition of an official U.S. Government military history classic provides an authoritative historical survey of the organization and accomplishments of the United States Army. This scholarly yet readable book is designed to inculcate an awareness of our nation's military past and to demonstrate that the study of military history is an essential ingredient in leadership development. It is also an essential addition to any personal military history library. This text is used in military ROTC training courses as a basic military history textbook. Volume 1 of 2 volume set.

Historically, nursing, in all of its missions of research/scholarship, education and practice, has not had access to large patient databases. Nursing consequently adopted qualitative methodologies with small sample sizes, clinical trials and lab research. Historically, large data methods were limited to traditional biostatistical analyses. In the United States, large payer data has been amassed and structures/organizations have been created to welcome scientists to explore these large data to advance knowledge discovery. Health systems electronic health records (EHRs) have now matured to generate massive databases with longitudinal trending. This text reflects how the learning health system infrastructure is maturing, and being advanced by health information exchanges (HIEs) with multiple organizations blending their data, or enabling distributed computing. It educates the readers on the evolution of knowledge discovery methods that span qualitative as well as quantitative data mining, including

Read Book Apache Spark Book Soup

the expanse of data visualization capacities, are enabling sophisticated discovery. New opportunities for nursing and call for new skills in research methodologies are being further enabled by new partnerships spanning all sectors.

In response to a recent surge of interest in Native American history, culture, and lore, Hippocrene brings you a concise and straightforward dictionary of the Navajo tongue. The dictionary is designed to aid Navajos learning English as well as English speakers interested in acquiring knowledge of Navajo. The largest of all the Native American tribes, the Navajo number about 125,000 and live mostly on reservations in Arizona, New Mexico, and Utah. Over 9,000 entries; A detailed section on Navajo pronunciation; A comprehensive, modern vocabulary; Useful, everyday expressions.

Explores the homogenization of American culture and the impact of the fast food industry on modern-day health, economy, politics, popular culture, entertainment, and food production.

The Visual Miscellaneum is a unique, groundbreaking look at the modern information age, helping readers make sense of the countless statistics and random facts that constantly bombard us. Using cutting edge graphs, charts, and illustrations, David McCandless creatively visualizes the world's surprising relationships and compelling data, covering everything from the most pleasurable guilty pleasures to how long it takes different condiments to spoil to world maps of Internet search terms.

This book describes the paleomagnetism of sediments and sedimentary rocks, how

Read Book Apache Spark Book Soup

sediments and sedimentary rocks become magnetized, and how the physical and chemical processes involved can affect the accuracy of paleomagnetism. Topics covered include depositional and post-depositional remanence acquisition, the detection and correction of compaction-caused inclination shallowing, reduction diagenesis of magnetic minerals, chemical remagnetization, and rotation of remanence by grain-scale rock strain. The book also has a chapter on environmental paleomagnetism, including examples of the new technique of high-resolution rock magnetic cyclostratigraphy and its application to sedimentary sequences. By emphasizing the accuracy of sedimentary paleomagnetism and the magnitude of post-depositional processes that can affect it, the book will be invaluable in the geologic interpretation of sedimentary paleomagnetic data. *Paleomagnetism of Sedimentary Rocks* will be welcomed by paleomagnetists, students of paleomagnetism and all Earth scientists who use sedimentary paleomagnetic data in their research. Additional resources for this book can be found at: www.wiley.com/go/kodama/paleomagnetism. Shortly after William Henry Fox Talbot announced his invention of photography in 1839, the dedicated amateur botanist Anna Atkins, daughter of a prominent British scientist, began to experiment with the new medium. In 1843 she turned to her friend Sir John Herschel's recently discovered cyanotype process to publish her growing collection of native seaweeds--a daring way to introduce photography into book illustration. At regular intervals over the next decade, Atkins printed and issued these bracingly

Read Book Apache Spark Book Soup

modern, deeply-hued photograms to her "botanical friends" in the form of hand-stitched fascicles of a book she entitled *Photographs of British Algæ: Cyanotype Impressions*. The first book to be illustrated by photography and the earliest sustained application of photography to science, *British Algæ* is a landmark in the histories of publishing and photography. Of the nearly two dozen substantially complete or partial copies known to exist, each is distinct in its appearance and often in its number and arrangement of plates. The set of 13 parts she gave to Sir John Herschel--now in the Spencer Collection of The New York Public Library--is especially important and was carefully preserved by generations of the Herschel family exactly as Sir John received it. This sumptuous facsimile edition reproduces the recto and verso of each plate, presenting the work as its creator intended: as bound volumes to lingered over, studied and admired, page by extraordinary page.

This book provides a complete and modern guide to web scraping, using Python as the programming language, without glossing over important details or best practices. Written with a data science audience in mind, the book explores both scraping and the larger context of web technologies in which it operates, to ensure full understanding. The authors recommend web scraping as a powerful tool for any data scientist's arsenal, as many data science projects start by obtaining an appropriate data set. Starting with a brief overview on scraping and real-life use cases, the authors explore the core concepts of HTTP, HTML, and CSS to provide a solid foundation. Along with a

Read Book Apache Spark Book Soup

quick Python primer, they cover Selenium for JavaScript-heavy sites, and web crawling in detail. The book finishes with a recap of best practices and a collection of examples that bring together everything you've learned and illustrate various data science use cases. What You'll Learn Leverage well-established best practices and commonly-used Python packages Handle today's web, including JavaScript, cookies, and common web scraping mitigation techniques Understand the managerial and legal concerns regarding web scraping Who This Book is For A data science oriented audience that is probably already familiar with Python or another programming language or analytical toolkit (R, SAS, SPSS, etc). Students or instructors in university courses may also benefit. Readers unfamiliar with Python will appreciate a quick Python primer in chapter 1 to catch up with the basics and provide pointers to other guides as well.

Summary Modern data science solutions need to be clean, easy to read, and scalable. In *Mastering Large Datasets with Python*, author J.T. Wolohan teaches you how to take a small project and scale it up using a functionally influenced approach to Python coding. You'll explore methods and built-in Python tools that lend themselves to clarity and scalability, like the high-performing parallelism method, as well as distributed technologies that allow for high data throughput. The abundant hands-on exercises in this practical tutorial will lock in these essential skills for any large-scale data science project. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the technology Programming techniques that

Read Book Apache Spark Book Soup

work well on laptop-sized data can slow to a crawl—or fail altogether—when applied to massive files or distributed datasets. By mastering the powerful map and reduce paradigm, along with the Python-based tools that support it, you can write data-centric applications that scale efficiently without requiring codebase rewrites as your requirements change. About the book *Mastering Large Datasets with Python* teaches you to write code that can handle datasets of any size. You'll start with laptop-sized datasets that teach you to parallelize data analysis by breaking large tasks into smaller ones that can run simultaneously. You'll then scale those same programs to industrial-sized datasets on a cluster of cloud servers. With the map and reduce paradigm firmly in place, you'll explore tools like Hadoop and PySpark to efficiently process massive distributed datasets, speed up decision-making with machine learning, and simplify your data storage with AWS S3. What's inside

An introduction to the map and reduce paradigm
Parallelization with the multiprocessing module and pathos framework
Hadoop and Spark for distributed computing
Running AWS jobs to process large datasets
About the reader
For Python programmers who need to work faster with more data.
About the author
J. T. Wolohan is a lead data scientist at Booz Allen Hamilton, and a PhD researcher at Indiana University, Bloomington.

Table of Contents: PART 1 1
! Introduction 2
! Accelerating large dataset work: Map and parallel computing 3
! Function pipelines for mapping complex transformations 4
! Processing large datasets with lazy workflows 5
! Accumulation operations with reduce 6
! Speeding up map and

reduce with advanced parallelization PART 2 7 | Processing truly big datasets with Hadoop and Spark 8 | Best practices for large data with Apache Streaming and mrjob 9 | PageRank with map and reduce in PySpark 10 | Faster decision-making with machine learning and PySpark PART 3 11 | Large datasets in the cloud with Amazon Web Services and S3 12 | MapReduce in the cloud with Amazon's Elastic MapReduce

How Boston radio station WBCN became the hub of the rock-and-roll, antiwar, psychedelic solar system. While San Francisco was celebrating a psychedelic Summer of Love in 1967, Boston stayed buttoned up and battened down. But that changed the following year, when a Harvard Law School graduate student named Ray Riepen founded a radio station that played music that young people, including the hundreds of thousands at Boston-area colleges, actually wanted to hear. WBCN-FM featured album cuts by such artists as the Mothers of Invention, Aretha Franklin, and Cream, played by announcers who felt free to express their opinions on subjects that ranged from recreational drugs to the war in Vietnam. In this engaging and generously illustrated chronicle, Peabody Award-winning journalist and one-time WBCN announcer Bill Lichtenstein tells the story of how a radio station became part of a revolution in youth culture. At WBCN, creativity and countercultural politics ruled: there were no set playlists; news segments anticipated the satire of The Daily Show; on-air interviewees ranged from John and Yoko to Noam Chomsky; a telephone "Listener Line" fielded questions on any subject, day and night. From 1968 to Watergate, Boston's WBCN

Read Book Apache Spark Book Soup

was the hub of the rock-and-roll, antiwar, psychedelic solar system. A cornucopia of images in color and black and white includes concert posters, news clippings, photographs of performers in action, and scenes of joyousness on Boston Common. Interwoven through the narrative are excerpts from interviews with WBCN pioneers, including Charles Laquidara, the “news dissector” Danny Schechter, Marsha Steinberg, and Mitchell Kertzman. Lichtenstein’s documentary WBCN and the American Revolution is available as a DVD sold separately.

The process of user-centered innovation: how it can benefit both users and manufacturers and how its emergence will bring changes in business models and in public policy. Innovation is rapidly becoming democratized. Users, aided by improvements in computer and communications technology, increasingly can develop their own new products and services. These innovating users—both individuals and firms—often freely share their innovations with others, creating user-innovation communities and a rich intellectual commons. In *Democratizing Innovation*, Eric von Hippel looks closely at this emerging system of user-centered innovation. He explains why and when users find it profitable to develop new products and services for themselves, and why it often pays users to reveal their innovations freely for the use of all. The trend toward democratized innovation can be seen in software and information products—most notably in the

Read Book Apache Spark Book Soup

free and open-source software movement—but also in physical products. Von Hippel's many examples of user innovation in action range from surgical equipment to surfboards to software security features. He shows that product and service development is concentrated among "lead users," who are ahead on marketplace trends and whose innovations are often commercially attractive. Von Hippel argues that manufacturers should redesign their innovation processes and that they should systematically seek out innovations developed by users. He points to businesses—the custom semiconductor industry is one example—that have learned to assist user-innovators by providing them with toolkits for developing new products. User innovation has a positive impact on social welfare, and von Hippel proposes that government policies, including R&D subsidies and tax credits, should be realigned to eliminate biases against it. The goal of a democratized user-centered innovation system, says von Hippel, is well worth striving for. An electronic version of this book is available under a Creative Commons license.

Red Sparrow is now a major motion picture starring Jennifer Lawrence and Joel Edgerton! The thrilling sequel to Red Sparrow—CIA insider Jason Matthews's compulsively readable New York Times bestseller and Edgar Award winner—featuring Russian spy Dominika Egorova and CIA agent Nate Nash

Read Book Apache Spark Book Soup

“shimmers with authenticity. The villains are richly drawn...the scenes of them on the job are beyond chilling” (The New York Times Book Review). Captain Dominika Egorova of the Russian Intelligence Service despises the oligarchs, crooks, and thugs of Putin’s Russia—but what no one knows is that she is also working for the CIA. Her “sparrow” training in the art of sexual espionage further complicates the mortal risks she must take, as does her love for her handler Nate Nash—a shared lust that is as dangerous as treason. As Dominika expertly dodges exposure, she deals with a murderously psychotic boss, survives an Iranian assassination attempt and attempts to rescue an arrested double agent—and thwart Putin’s threatening flirtations. A grand, wildly entertaining ride through the steel-trap mind of a CIA insider, *Palace of Treason* is a story “as suspenseful and cinematic as the best spy movies” (The Philadelphia Inquirer)—one that feels fresh and so possible, in fact, that it’s doubtful this novel can ever be published in Russia.

'Things Fall Apart' tells the story of Okonkwo, an important man in the Igbo tribe in the days when white men were first on the scene. Okonkwo becomes exiled from his tribe, as a result of his pride and his fears, with tragic consequences. This is a book about the big questions in life: knowledge, consciousness, fate, God, truth, goodness, justice. It is for anyone who believes there are big

Read Book Apache Spark Book Soup

questions out there, but does not know how to approach them. Think sets out to explain what they are and why they are important. Simon Blackburn begins by putting forward a convincing case for the study of philosophy and goes on to give the reader a sense of how the great historical figures such as Descartes, Hume, Kant, and Wittgenstein have approached its central themes. Each chapter explains a major issue, and gives the reader a self-contained guide through the problems that philosophers have studied. The large scope of topics covered range from scepticism, the self, mind and body, and freedom to ethics and the arguments surrounding the existence of God. Lively and approachable, this book is ideal for all those who want to learn how the basic techniques of thinking shape our existence.

Getting Started with Beautiful SoupPackt Publishing Ltd

A sampling of the writings of Chinese women poets from ancient times to the twentieth century is accompanied by brief biographies and a discussion of their influence on Chinese literature.

Open the pages of this lively book and discover a bowl of animal alphabet soup sure to spark children's imaginations. From the tiny ant on the front endpapers to the exuberant zebra in the final spread, children will delight in the mischievous menagerie that bursts onto the table, while at the same time learning their letters

Read Book Apache Spark Book Soup

and animals. Filled with imaginative details, it is a visual feast from start to finish. Build, process and analyze large-scale graph data effectively with Spark About This Book Find solutions for every stage of data processing from loading and transforming graph data to Improve the scalability of your graphs with a variety of real-world applications with complete Scala code. A concise guide to processing large-scale networks with Apache Spark. Who This Book Is For This book is for data scientists and big data developers who want to learn the processing and analyzing graph datasets at scale. Basic programming experience with Scala is assumed. Basic knowledge of Spark is assumed. What You Will Learn Write, build and deploy Spark applications with the Scala Build Tool. Build and analyze large-scale network datasets Analyze and transform graphs using RDD and graph-specific operations Implement new custom graph operations tailored to specific needs. Develop iterative and efficient graph algorithms using message aggregation and Pregel abstraction Extract subgraphs and use it to discover common clusters Analyze graph data and solve various data science problems using real-world datasets. In Detail Apache Spark is the next standard of open-source cluster-computing engine for processing big data. Many practical computing problems concern large graphs, like the Web graph and various social networks. The scale of these graphs - in some cases billions of vertices, trillions

of edges - poses challenges to their efficient processing. Apache Spark GraphX API combines the advantages of both data-parallel and graph-parallel systems by efficiently expressing graph computation within the Spark data-parallel framework. This book will teach the user to do graphical programming in Apache Spark, apart from an explanation of the entire process of graphical data analysis. You will journey through the creation of graphs, its uses, its exploration and analysis and finally will also cover the conversion of graph elements into graph structures. This book begins with an introduction of the Spark system, its libraries and the Scala Build Tool. Using a hands-on approach, this book will quickly teach you how to install and leverage Spark interactively on the command line and in a standalone Scala program. Then, it presents all the methods for building Spark graphs using illustrative network datasets. Next, it will walk you through the process of exploring, visualizing and analyzing different network characteristics. This book will also teach you how to transform raw datasets into a usable form. In addition, you will learn powerful operations that can be used to transform graph elements and graph structures. Furthermore, this book also teaches how to create custom graph operations that are tailored for specific needs with efficiency in mind. The later chapters of this book cover more advanced topics such as clustering graphs, implementing graph-parallel iterative algorithms and learning

Read Book Apache Spark Book Soup

methods from graph data. Style and approach A step-by-step guide that will walk you through the key ideas and techniques for processing big graph data at scale, with practical examples that will ensure an overall understanding of the concepts of Spark.

[Copyright: 03c62e07c84dd34b92195335c4462fbf](https://www.dbooks.org/doc/03c62e07c84dd34b92195335c4462fbf)