Building Search Applications Lucene Lingpipe And Gate

Secondary Use of Electronic Patient Records
Version 6
The 8051/8052 Microcontroller
High Frequency MOSFET Gate Drivers
Natural Language Processing with Java
Techniques for building machine learning and neural network models for NLP, 2nd Edition
Lucene, LingPipe, and Gate
From Theory to Practice
9th International Conference, SEAL 2012, Hanoi, Vietnam, December 16-19, 2012, Proceedings
Proceedings of the Second International Afro-European Conference for Industrial Advancement AECIA 2015
9th International Conference, SpaCCS 2016, Zhangjiajie, China, November 16-18, 2016, Proceedings
Developing Large Web Applications
Introduction to Linguistic Annotation and Text Analytics
Information Systems: Crossroads for Organization, Management, Accounting and Engineering
Architecture, Assembly Language, and Hardware Interfacing
Building Search Applications with Lucene and Lingpipe
Text Processing in Java
Designing and Building Big Data Systems using the Hadoop Ecosystem
Lucene in Action
Second International Conference, EGOVIS 2011, Toulouse, France, August 29 -- September 2, 2011, Proceedings
Mobile Computing and Wireless Communications
Machine Learning and Knowledge Discovery in Databases
Knowledge Technology
A Cognitive Approach Based on Natural Language Processing
International Workshops of ECML PKDD 2019, Würzburg, Germany, September 16–20, 2019, Proceedings, Part II
Against a Sharp White Background
Ancient Greek and Latin in the Digital Revolution
Biomedical Natural Language Processing
Essential GWT
Optimizing Human-Computer Interaction With Emerging Technologies
Clinical Text Mining
Security, Privacy, and Anonymity in Computation, Communication, and Storage
Text Mining Application Programming

Third Knowledge Technology Week, KTW 2011, Kajang, Malaysia, July 18-22, 2011. Revised Selected Papers
Visual Information Communication
ItAIS: The Italian Association for Information Systems
Pro Hadoop Data Analytics
Electronic Government and the Information Systems Perspective
Applications, Networks, Platforms, Architectures, and Security
Mastering Java for Data Science

MARSHALL ANNABEL

Secondary Use of Electronic Patient Records
Springer Science & Business Media
When Lucene first hit the scene five years ago, it was nothing short ofamazing. By using this open-source, highly scalable, super-fast search engine,developers could integrate search into applications quickly and efficiently.A lot has changed since then-search has grown from a "nice-to-have" featureinto an indispensable part of most enterprise applications. Lucene now powerssearch in diverse companies including Akamai, Netflix, LinkedIn,Technorati, HotJobs, Epiphany, FedEx, Mayo Clinic, MIT, New ScientistMagazine, and many others. Some things remain the same, though. Lucene still delivers high-performancesearch features in a disarmingly easy-to-use API. Due to its vibrant and diverseopen-source community of developers and users, Lucene is relentlessly improving,with evolutions to APIs, significant new features such as payloads, and ahuge increase (as much as 8x) in indexing speed with Lucene 2.3. And with clear writing, reusable examples, and unmatched advice on bestpractices, Lucene in Action, Second Edition is still the definitive guide todeveloping with Lucene. 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.
Version 6 Packt Publishing Ltd
This book teaches you how to master the subtle art of multilingual text processing and prevent text data corruption. It provides an introduction to natural language processing using Lucene and Solr. It gives you tools and techniques to manage large collections of text data, whether they come from news feeds, databases, or legacy documents. Each chapter contains executable programs that can also be used for text data forensics. Topics covered: Unicode code points Character encodings from ASCII and Big5 to UTF-8 and UTF-32LE Character normalization using International Components for Unicode (ICU) Java I/O, including working directly with zip, gzip, and tar files Regular expressions in Java Transporting text data via HTTP Parsing and generating XML, HTML, and JSON Using Lucene 4 for natural language search and text classification Search, spelling correction, and clustering with Solr 4 Other books on text processing presuppose much of the material covered in this book. They gloss over the details of transforming text from one format to another and

assume perfect input data. The messy reality of raw text will have you reaching for this book again and again.
The 8051/8052 Microcontroller IBM Press
This book was written with the novice or intermediate 8052 developer in mind. Assuming no prior knowledge of the 8052, it takes the reader step-by-step through the architecture including discussions and explanations of concepts such as internal RAM, external RAM, Special Function Registers (SFRs), addressing modes, timers, serial I/O, and interrupts. This is followed by an in-depth section on assembly language which explains each instruction in the 8052 instruction set as well as related concepts such as assembly language syntax, expressions, assembly language directives, and how to implement 16-bit mathematical functions. The book continues with a thorough explanation of the 8052 hardware itself, reviewing the function of each pin on the microcontroller and follows this with the design and explanation of a fully functional single board computer-every section of the schematic design is explained in detail to provide the reader with a full understanding of how everything is connected, and why. The book closes with a section on hardware interfacing and software examples in which the reader will learn about the SBCMON monitor program for use on the single board computer, interfacing with a 4x4 keypad, communicating with a 16x2 LCD in direct-connect as well as memory-mapped fashion, utilizing an external serial EEPROM via the SPI protocol, and using the I2C communication standard to access an external real time clock. The book takes the reader with absolutely no knowledge of the 8052 and provides him with the information necessary to understand the architecture, design and build a functioning circuit based on the 8052, and write software to operate the 8052 in assembly language.

High Frequency MOSFET Gate Drivers

Building Search Applications with Lucene and LingpipeBuilding Search ApplicationsLucene, LingPipe, and Gate
This book examines a wide range of issues that characterize the current IT based innovation trends in organizations. It contains a collection of research papers focusing on themes of growing interest in the field of Information Systems, Organization Studies, Management, Accounting and Engineering. The book offers a multidisciplinary view on Information Systems with the aim of disseminating academic knowledge. It would be particularly relevant to IT practitioners such as information systems managers and IT consultants. The 12 sections cover a broad spectrum of topics including: eServices in Public and Private Sectors; Organizational Change and the Impact of ICT in Public and Private Sectors; Information and Knowledge Management; Human-Computer Interaction; Information Systems, Innovation Transfer, and New Business Models; Business Intelligence Systems, their Strategic Role and Organizational Impacts; New Ways to Work and Interact with the Internet; IS, IT and Security; Blending Design and Behavioral Research in Information Systems; Professional Skills,

Certification of Curricula, Online Education and Communities; IS Design, IS Development, Metrics and Compliance; ICT4LAW: Information and communication technologies to help firms, public administrations, legislators and citizens to operate in a highly regulated world. The content of each section is based on a selection of original double-blind peer reviewed contributions.

Natural Language Processing with Java

Springer

Visual communication through graphical and sign languages has long been conducted among human beings of different backgrounds and cultures, and in recent decades between human and machine. In today's digital world, visual information is typically encoded with various metaphors commonly used in daily life to facilitate rapid comprehension and easy analysis during the communication process. Visual information communication generally encompasses information visualization, graphical user-interfaces, visual analytics, visual languages and multi-media processing. It has been successfully employed in knowledge discovery, end-user programming, modeling, rapid systems prototyping, education, and design activities by people of many disciplines including architects, artists, children, engineers, and scientists. In addition, visual information is increasingly being used to facilitate human-human communication through the Internet and Web technology, and electronic mobile devices. This manuscript provides the cutting-edge techniques, approaches and the latest ongoing researches in the context of visual information communication. It is a collection of 24 chapters selected from more than 60 submissions to the VINCI'09 - 2009 Visual Information Communications International Conference, that is held in Sydney Australia, September 2009. These chapters were selected through a stringent review process to ensure their high standard in quality, significance and relevance. Each chapter was reviewed by at least two international Program Committee members of VINCI'09. The book covers a broad range of contents in five key sub-areas of visual information communication, including.

Techniques for building machine learning and neural network models for NLP, 2nd Edition

Lulu.com

Information retrieval systems centrally build upon the concept of relevance in order to rank documents in response to a user's query. Assessing relevance is a non-trivial operation that can be influenced by a multitude of factors that go beyond mere topical overlap with the query. This thesis argues that relevance depends on personal (Chapter 2) and situational (Chapter 3) context. In many use cases, there is no single interpretation of the concept that would optimally satisfy all users in all possible situations.We postulate that relevance should be explicitly modelled as a composite notion comprised of individual relevance dimensions. To this end, we show how automatic inference schemes based on document content and user activity can be used in order to estimate such constituents of relevance (Chapter 4). Alternatively, we can employ human expertise, harnessed, for example, via commercial

crowdsourcing or serious games to judge the degree to which a document satisfies a given set of relevance dimensions (Chapter 5).Finally, we need a model that allows us to estimate the joint distribution of relevance across all previously obtained dimensions. In this thesis, we propose using copulas, a model family originating from the field of quantitative finances that decouples observations and dependency structure and which can account for complex non-linear dependencies among relevance dimensions (Chapter 6).

Lucene, LingPipe, and Gate Springer

Lucene, LingPipe, and Gate are popular open source tools to build powerful search applications. Building Search Applications describes functions from Lucene that include indexing, searching, ranking, and spelling correction to build search engines. With this book you will learn to: Extract tokens from text using custom tokenizers and analyzers from Lucene, LingPipe, and Gate. Construct a search engine index with an optional backend database to manage large document collections. Explore the wide range of Lucene queries to search an index, understand the ranking algorithm for a query, and suggest spelling corrections. Find the names of people, places, and other entities in text using LingPipe and Gate. Categorize documents by topic using classifiers and build groups of self-organized documents using clustering algorithms from LingPipe. Create a Web crawler to scan the Web, Intranet, or desktop using Nutch. Track the sentiment of articles published on the Web with LingPipe.

From Theory to Practice CRC Press

This book, suitable for IS/IT courses and self study, presents a comprehensive coverage of the technical as well as business/management aspects of mobile computing and wireless communications. Instead of one narrow topic, this classroom tested book covers the major building blocks (mobile applications, mobile computing platforms, wireless networks, architectures, security, and management) of mobile computing and wireless communications. Numerous real-life case studies and examples highlight the key points. The book starts with a discussion of m-business and m-government initiatives and examines mobile computing applications such as mobile messaging, m-commerce, M-CRM, M-portals, M-SCM, mobile agents, and sensor applications. The role of wireless Internet and Mobile IP is explained and the mobile computing platforms are analyzed with a discussion of wireless middleware, wireless gateways, mobile application servers, WAP, i-mode, J2ME, BREW, Mobile Internet Toolkit, and Mobile Web Services. The wireless networks are discussed at length with a review of wireless communication principles, wireless LANs with emphasis on 802.11 LANs, Bluetooth, wireless sensor networks, UWB (Ultra Wideband), cellular networks ranging from 1G to 5G, wireless local loops, FSO (Free Space Optics), satellites communications, and deep space networks. The book concludes with a review of the architectural, security, and management/support issues and their role in building, deploying and

managing wireless systems in modern settings.
9th International Conference, SEAL 2012, Hanoi, Vietnam, December 16-19, 2012, Proceedings Packt Publishing Ltd
This two-volume set constitutes the refereed proceedings of the workshops which complemented the 19th Joint European Conference on Machine Learning and Knowledge Discovery in Databases, ECML PKDD, held in Würzburg, Germany, in September 2019. The 70 full papers and 46 short papers presented in the two-volume set were carefully reviewed and selected from 200 submissions. The two volumes (CCIS 1167 and CCIS 1168) present the papers that have been accepted for the following workshops: Workshop on Automating Data Science, ADS 2019; Workshop on Advances in Interpretable Machine Learning and Artificial Intelligence and eXplainable Knowledge Discovery in Data Mining, AIMLAI-XKDD 2019; Workshop on Decentralized Machine Learning at the Edge, DMLE 2019; Workshop on Advances in Managing and Mining Large Evolving Graphs, LEG 2019; Workshop on Data and Machine Learning Advances with Multiple Views; Workshop on New Trends in Representation Learning with Knowledge Graphs; Workshop on Data Science for Social Good, SoGood 2019; Workshop on Knowledge Discovery and User Modelling for Smart Cities, UMCIT 2019; Workshop on Data Integration and Applications Workshop, DINA 2019; Workshop on Machine Learning for Cybersecurity, MLCS 2019; Workshop on Sports Analytics: Machine Learning and Data Mining for Sports Analytics, MLSA 2019; Workshop on Categorising Different Types of Online Harassment Languages in Social Media; Workshop on IoT Stream for Data Driven Predictive Maintenance, IoTStream 2019; Workshop on Machine Learning and Music, MML 2019; Workshop on Large-Scale Biomedical Semantic Indexing and Question Answering, BioASQ 2019.
Proceedings of the Second International Afro-European Conference for Industrial Advancement AECIA 2015 Springer Nature
The second edition of this book builds all the code example within a single project by incorporating new advancements in C# .NET technology and open-source math libraries. It also uses C# Interactive Window to test numerical computations without compiling or running the complete project code. The second edition includes three new chapters, including "Plotting", Fourier Analysis" and "Math Expression Parser". As in the first edition, this book presents an in-depth exposition of the various numerical methods used in real-world scientific and engineering computations. It emphasizes the practical aspects of C# numerical methods and mathematical functions programming, and discusses various techniques in details to enable you to implement these numerical methods in your .NET application. Ideal for scientists, engineers, and students who would like to become more adept at numerical methods, the second edition of this book covers the following content: - Overview of C# programming. - The mathematical background and fundamentals of numerical methods. - plotting the computation

results using a 3D chart control. - Math libraries for complex numbers and functions, real and complex vector and matrix operations, and special functions. - Numerical methods for generating random numbers and random distribution functions. - Various numerical methods for solving linear and nonlinear equations. - Numerical differentiation and integration. - Interpolations and curve fitting. - Optimization of single-variable and multi-variable functions with a variety of techniques, including advanced simulated annealing and evolutionary algorithms. - Numerical techniques for solving ordinary differential equations. - Numerical methods for solving boundary value problems. - Eigenvalue problems. - Fourier analysis. - mathematical expression parser and evaluator. In addition, this book provides testing examples for every math function and numerical method to show you how to use these functions and methods in your own .NET applications in a manageable and step-by-step fashion. Please visit the author's website for more information about this book at https://drxudotnet.com https://drxudotnet.com and https://gincker.com.

9th International Conference, SpaCCS 2016, Zhangjiajie, China, November 16-18, 2016, Proceedings Pearson Education

Biological and biomedical research are increasingly driven by experimental techniques that challenge our ability to analyse, process and extract meaningful knowledge from the underlying data. The impressive capabilities of next generation sequencing technologies, together with novel and ever evolving distinct types of omics data technologies, have put an increasingly complex set of challenges for the growing fields of Bioinformatics and Computational Biology. The analysis of the datasets produced and their integration call for new algorithms and approaches from fields such as Databases, Statistics, Data Mining, Machine Learning, Optimization, Computer Science and Artificial Intelligence. Clearly, Biology is more and more a science of information requiring tools from the computational sciences. In the last few years, we have seen the surge of a new generation of interdisciplinary scientists that have a strong background in the biological and computational sciences. In this context, the interaction of researchers from different scientific fields is, more than ever, of foremost importance boosting the research efforts in the field and contributing to the education of a new generation of Bioinformatics scientists. PACBB'14 contributes to this effort promoting this fruitful interaction. PACBB'14 technical program included 34 papers spanning many different sub-fields in Bioinformatics and Computational Biology. Therefore, the conference promotes the interaction of scientists from diverse research groups and with a distinct background such as computer scientists, mathematicians or biologists.

Developing Large Web Applications University of Wisconsin Press

Die verbreiteten Begriffe ‚Informationsgesellschaft' und ‚Age of Access' suggerieren die problemlose allseitige Zugänglichkeit von Information. Doch Information ist in der

Realität in vielerlei Hinsicht unzugänglich - physisch, wirtschaftlich, intellektuell, sprachlich, politisch, technisch. Zudem entstehen täglich neue Techniken und Praktiken der Zugänglichmachung. Schließlich zeigen sich in verschiedenen Bereichen die Grenzen der Forderung nach Zugänglichkeit. Diese neue Buchreihe bringt Wissenschaftler und Praktiker verschiedenster Prägung zusammen, um die verschiedenen Dimensionen der Unzugänglichkeit von Information auszuloten sowie Prinzipien und Techniken ihrer praktischen und gesellschaftlichen Überwindung aufzuzeigen, aber auch notwendige Grenzen der Zugänglichkeit deutlich zu machen. Herausgegeben von André Schüller-Zwierlein, Universitätsbibliothek Regensburg. Editorial Board: Prof. Dr. Herbert Burkert (Informationsrecht, Universität St. Gallen) Dr. Klaus Ceynowa (Stv. Generaldirektor der Bayerischen Staatsbibliothek) Prof. Dr. Heinrich Hußmann (Angewandte Informatik und Medieninformatik, Ludwig-Maximilians-Universität München) Prof. Dr. Michael Jäckel (Soziologie, Universität Trier) Prof. Dr. Rainer Kuhlen (Informationswissenschaft , Universität Konstanz) Prof. Dr. Frank Marcinkowski (Kommunikationswissenschaft, Westfälische Wilhelms-Universität Münster) Prof. Dr. Rudi Schmiede (Soziologie, Technische Universität Darmstadt) Prof. Dr. Richard Stang (Bibliotheks- und Informationsmanagement, Hochschule der Medien, Stuttgart)

Introduction to Linguistic Annotation and Text Analytics Morgan & Claypool Publishers

Text Mining Application Programming teaches software developers how to mine the vast amounts of information available on the Web, internal networks, and desktop files and turn it into usable data. The book helps developers understand the problems associated with managing unstructured text, and explains how to build your own mining tools using standard statistical methods from information theory, artificial intelligence, and operations research. Each of the topics covered are thoroughly explained and then a practical implementation is provided.The book begins with a brief overview of text data, where it can be found, and the typical search engines and tools used to search and gather this text. It details how to build tools for extracting and using the text, and covers the mathematics behind many of the algorithms used in building these tools. From there you'll learn how to build tokens from text, construct indexes, and detect patterns in text. You'll also find methods to extract the names of people, places, and organizations from an email, a news article, or a Web page. The next portion of the book teaches you how to find information on the Web, the structure of the Web, and how to build spiders to crawl the Web. Text categorization is also described in the context of managing email. The final part of the book covers information monitoring, summarization, and a simple Question & Answer (Q&A) system. The code used in the book is written in Perl, but knowledge of Perl is not necessary to run the software.

Developers with an intermediate level of experience with Perl can customize the software. Although the book is about programming, methods are explained with English-like pseudocode and the source code is provided on the CD-ROM. After reading this book, you'll be ready to tap into the bevy of information available online in ways you never thought possible.

Information Systems: Crossroads for Organization, Management, Accounting and Engineering Packt Publishing Ltd

This book constitutes the refereed proceedings of the 9th International Conference on on Security, Privacy and Anonymity in Computation, Communication and Storage, SpaCCS 2016, held in Zhangjiajie, China, in November 2016. The 40 papers presented in this volume were carefully reviewed and selected from 110 submissions. They are organized in topical sections including security algorithms and architectures, privacy-aware policies, regulations and techniques, anonymous computation and communication, encompassing fundamental theoretical approaches, practical experimental projects, and commercial application systems for computation, communication and storage.

Architecture, Assembly Language, and Hardware Interfacing IGI Global

Learn advanced analytical techniques and leverage existing tool kits to make your analytic applications more powerful, precise, and efficient. This book provides the right combination of architecture, design, and implementation information to create analytical systems that go beyond the basics of classification, clustering, and recommendation. Pro Hadoop Data Analytics emphasizes best practices to ensure coherent, efficient development. A complete example system will be developed using standard third-party components that consist of the tool kits, libraries, visualization and reporting code, as well as support glue to provide a working and extensible end-to-end system. The book also highlights the importance of end-to-end, flexible, configurable, high-performance data pipeline systems with analytical components as well as appropriate visualization results. You'll discover the importance of mix-and-match or hybrid systems, using different analytical components in one application. This hybrid approach will be prominent in the examples. What You'll Learn Build big data analytic systems with the Hadoop ecosystem Use libraries, tool kits, and algorithms to make development easier and more effective Apply metrics to measure performance and efficiency of components and systems Connect to standard relational databases, noSQL data sources, and more Follow case studies with example components to create your own systems Who This Book Is For Software engineers, architects, and data scientists with an interest in the design and implementation of big data analytical systems using Hadoop, the Hadoop ecosystem, and other associated technologies.

Building Search Applications with Lucene and Lingpipe Springer

This book constitutes the refereed proceedings of the Third Knowledge Technology Week, KTW

2011, held in Kajang, Malaysia, in July 2011. The 29 revised full papers presented together with 9 short papers were carefully reviewed and selected from 105 submissions. KTW 2011 consisted of a number of co-located events. This volume contains selected papers from the proceedings of the Third Malaysian Joint Conference on Artificial Intelligence (MJCAI 2011), the Third Semantic Technology and Knowledge Engineering (STAKE 2011), and the International Workshop on Semantic Agents (IWSA 2012).

Text Processing in Java

Createspace Independent Pub

This book describes high frequency power MOSFET gate driver technologies, including gate drivers for GaN HEMTs, which have great potential in the next generation of switching power converters. Gate drivers serve as a critical role between control and power devices. In recent years, there has been a trend to increase the switching frequency beyond multi-MHz in switching power converters to reduce the passive components and significantly improve power density. However, this results in high switching loss and gate driver loss in power MOSFETs. The novel approach in this book is the proposed Current Source Gate Driver (CSD) including different topologies, control and applications. The CSD can reduce the switching transition time and switching loss significantly, and recover high frequency gate driver loss compared to conventional voltage gate drivers. The basic idea can also be extended to other power devices to improve high frequency switching performance such as SiC MOSFET and IGBT. Topics covered in the book include the state-of-the-art of power MOSFET drive techniques, the switching loss model, current source gate drivers (CSDs), resonant gate drivers, adaptive gate drivers and GaN HEMT gate drivers. The book is essential reading for design engineers, researchers and advanced students working in switching power supplies and in power electronics generally.

Designing and Building Big Data Systems using the Hadoop Ecosystem

nge solutions, inc

This open access book describes the results of natural language processing and machine learning methods applied to clinical text from electronic patient records. It is divided into twelve chapters. Chapters 1-4 discuss the history and background of the original paper-based patient records, their purpose, and how they are written and structured. These initial chapters do not require any technical or medical background knowledge. The remaining eight chapters are more technical in nature and describe various medical classifications and terminologies such as ICD diagnosis codes, SNOMED CT, MeSH, UMLS, and ATC. Chapters 5-10 cover basic tools for natural language processing and information retrieval, and how to apply them to clinical text. The difference between rule-based and machine learning-based methods, as well as between supervised and unsupervised machine learning methods, are also explained. Next, ethical concerns regarding the use of sensitive patient records for research purposes are discussed, including methods for de-identifying electronic patient records and safely storing patient

records. The book’s closing chapters present a number of applications in clinical text mining and summarise the lessons learned from the previous chapters. The book provides a comprehensive overview of technical issues arising in clinical text mining, and offers a valuable guide for advanced students in health informatics, computational linguistics, and information retrieval, and for researchers entering these fields.

Lucene in Action Esri Press

With the advent and increasing popularity of Computer Supported Collaborative Learning (CSCL) and e-learning technologies, the need of automatic assessment and of teacher/tutor support for the two tightly intertwined activities of comprehension of reading materials and of collaboration among peers has grown significantly. In this context, a polyphonic model of discourse derived from Bakhtin’s work as a paradigm is used for analyzing both general texts and CSCL conversations in a unique framework focused on different facets of textual cohesion. As specificity of our analysis, the individual learning perspective is focused on the identification of reading strategies and on providing a multi-dimensional textual complexity model, whereas the collaborative learning dimension is centered on the evaluation of participants’ involvement, as well as on collaboration assessment. Our approach based on advanced Natural Language Processing techniques provides a qualitative estimation of the learning process and enhances understanding as a “mediator of learning” by providing automated feedback to both learners and teachers or tutors. The main benefits are its flexibility, extensibility and nevertheless specificity for covering multiple stages, starting from reading classroom materials, to discussing on specific topics in a collaborative manner and finishing the feedback loop by verbalizing metacognitive thoughts.

Second International Conference, EGOVIS 2011, Toulouse, France, August 29 -- September 2, 2011, Proceedings Springer

Become an expert at building and deploying enterprise-grade data applications in JavaAbout This Book* This comprehensive book shows you exactly how you can take your Java data science applications to production seamlessly* Dive deep into analytics, supervised and unsupervised learning, and much more with ease* Explore Java's various libraries to efficiently build and deploy data applications for the enterpriseWho This Book Is ForThis book is for those Java developers who are comfortable with developing applications in Java and are familiar with the basic concepts of data science. This is the go-to book for anyone looking to master the subject using Java. If you are willing to build efficient data applications in your enterprise environment without changing your existing stack, this book is for you!What you will learn* Get a solid understanding of the data processing toolbox available in Java* Explore the data science ecosystem available in Java and other JVM languages* Understand when to use Java and what is best to do outside of Java* Deal with the machine learning task at hand and bring the results directly to production*

Get state-of-the-art performance with xgboost and deeplearning4j* Build applications that scale and process large amounts of data in real timeIn DetailJava is the language of choice if you want to bring data science to production, thanks to its stability and rich set of libraries. Major big data solutions including Hadoop are written in Java. This book will teach you how to perform data analysis on big data in a much more sophisticated manner. If you are willing to take your data products to enterprise without changing your stack, this book will tell you how to do it with ease.This book will quickly brush up on what you already know about using Java in data science applications and will then dive quickly into the advanced concepts to implement data science in production. The book covers topics such as advanced data science algorithms, preparing tricky data, advanced clustering, regression, classification, prediction, machine learning, and more.We'll teach you how data science can be used effectively to analyze unstructured data and big data. This book will enable you to tackle the problems of advanced visualization, advanced statistics, scaling data science applications, deploying these applications in production, and many more. You will also learn about natural language processing, real-time analytics, deep learning, and neural networks.