

Text Processing In Java

bull; Demonstrates how Python is the perfect language for text-processing functions. bull; Provides practical pointers and tips that emphasize efficient, flexible, and maintainable approaches to text-processing challenges. bull; Helps programmers develop solutions for dealing with the increasing amounts of data with which we are all inundated.

Text is everywhere. Web pages, databases, the contents of files--for almost any programming task you perform, you need to process text. Cut even the most complex text-based tasks down to size and learn how to master regular expressions, scrape information from Web pages, develop reusable utilities to process text in pipelines, and more. Most information in the world is in text format, and programmers often find themselves needing to make sense of the data hiding within. It might be to convert it from one format to another, or to find out information about the text as a whole, or to extract information from it. But how do you do this efficiently, avoiding labor-intensive, manual work? Text Processing with Ruby takes a practical approach. You'll learn how to get text into your Ruby programs from the file system and from user input. You'll process delimited files such as CSVs, and write utilities that interact with other programs in text-processing pipelines. Decipher character encoding mysteries, and avoid the pain of jumbled characters and malformed output. You'll learn to use regular expressions to match, extract, and replace patterns in text. You'll write a parser and learn how to process Web pages to pull out information from even the messiest of HTML. Before long you'll be able to tackle even the most enormous and entangled text with ease, scything through gigabytes of data and effortlessly extracting the bits that matter. What You Need: This book requires a passing familiarity with the Ruby programming language, and assumes that you already have Ruby installed on your computer.

While most other image processing texts approach this subject from an engineering perspective, The Art of Image Processing with Java places image processing within the realm of both engineering and computer science students by emphasizing software design. Ideal for students studying computer science or software engineering, it clearly teaches

Much of the data available today is unstructured and text-heavy, making it challenging for analysts to apply their usual data wrangling and visualization tools. With this practical book, you'll explore text-mining techniques with tidytext, a package that authors Julia Silge and David Robinson developed using the tidy principles behind R packages like ggraph and dplyr. You'll learn how tidytext and other tidy tools in R can make text analysis easier and more effective. The authors demonstrate how treating text as data frames enables you to manipulate, summarize, and visualize characteristics of text. You'll also learn how to integrate natural language processing (NLP) into effective workflows. Practical code examples and data explorations will help you generate real insights from literature, news, and social media. Learn how to apply the tidy text format to NLP Use sentiment analysis to mine the emotional content of text Identify a document's most important terms with frequency measurements Explore relationships and connections between words with the ggraph and widyr packages Convert back and forth between R's tidy and non-tidy text formats Use topic modeling to classify document collections into natural groups Examine case studies that compare Twitter archives, dig into NASA metadata, and analyze thousands of Usenet messages

Use Java and Deeplearning4j to build robust, scalable, and highly accurate AI models from scratch Key Features Install and configure Deeplearning4j to implement deep learning models from scratch Explore recipes for developing, training, and fine-tuning your neural network models in Java Model neural networks using datasets containing images, text, and time-series data Book Description Java is one of the most widely used programming languages in the world. With this book, you will see how to perform deep learning using Deeplearning4j (DL4J) – the most popular Java library for training neural networks efficiently. This book starts by

showing you how to install and configure Java and DL4J on your system. You will then gain insights into deep learning basics and use your knowledge to create a deep neural network for binary classification from scratch. As you progress, you will discover how to build a convolutional neural network (CNN) in DL4J, and understand how to construct numeric vectors from text. This deep learning book will also guide you through performing anomaly detection on unsupervised data and help you set up neural networks in distributed systems effectively. In addition to this, you will learn how to import models from Keras and change the configuration in a pre-trained DL4J model. Finally, you will explore benchmarking in DL4J and optimize neural networks for optimal results. By the end of this book, you will have a clear understanding of how you can use DL4J to build robust deep learning applications in Java. What you will learn

- Perform data normalization and wrangling using DL4J
- Build deep neural networks using DL4J
- Implement CNNs to solve image classification problems
- Train autoencoders to solve anomaly detection problems using DL4J
- Perform benchmarking and optimization to improve your model's performance
- Implement reinforcement learning for real-world use cases using RL4J
- Leverage the capabilities of DL4J in distributed systems

Who this book is for If you are a data scientist, machine learning developer, or a deep learning enthusiast who wants to implement deep learning models in Java, this book is for you. Basic understanding of Java programming as well as some experience with machine learning and neural networks is required to get the most out of this book.

If you want to build an enterprise-quality application that uses natural language text but aren't sure where to begin or what tools to use, this practical guide will help get you started. Alex Thomas, principal data scientist at Wisecube, shows software engineers and data scientists how to build scalable natural language processing (NLP) applications using deep learning and the Apache Spark NLP library. Through concrete examples, practical and theoretical explanations, and hands-on exercises for using NLP on the Spark processing framework, this book teaches you everything from basic linguistics and writing systems to sentiment analysis and search engines. You'll also explore special concerns for developing text-based applications, such as performance. In four sections, you'll learn NLP basics and building blocks before diving into application and system building:

- Basics:** Understand the fundamentals of natural language processing, NLP on Apache Spark, and deep learning
- Building blocks:** Learn techniques for building NLP applications—including tokenization, sentence segmentation, and named-entity recognition—and discover how and why they work
- Applications:** Explore the design, development, and experimentation process for building your own NLP applications
- Building NLP systems:** Consider options for productionizing and deploying NLP models, including which human languages to support

In recent years, online social networking has revolutionized interpersonal communication. The newer research on language analysis in social media has been increasingly focusing on the latter's impact on our daily lives, both on a personal and a professional level. Natural language processing (NLP) is one of the most promising avenues for social media data processing. It is a scientific challenge to develop powerful methods and algorithms which extract relevant information from a large volume of data coming from multiple sources and languages in various formats or in free form. We discuss the challenges in analyzing social media texts in contrast with traditional documents. Research methods in information extraction, automatic categorization and clustering, automatic summarization and indexing, and statistical machine translation need to be adapted to a new kind of data. This book reviews the current research on NLP tools and methods for processing the non-traditional information from social media data that is available in large amounts (big data), and shows how innovative NLP approaches can integrate appropriate linguistic information in various fields such as social media monitoring, healthcare, business intelligence, industry, marketing, and security and defence. We review the existing evaluation metrics for NLP and social media applications, and the new

efforts in evaluation campaigns or shared tasks on new datasets collected from social media. Such tasks are organized by the Association for Computational Linguistics (such as SemEval tasks) or by the National Institute of Standards and Technology via the Text REtrieval Conference (TREC) and the Text Analysis Conference (TAC). In the concluding chapter, we discuss the importance of this dynamic discipline and its great potential for NLP in the coming decade, in the context of changes in mobile technology, cloud computing, virtual reality, and social networking. In this second edition, we have added information about recent progress in the tasks and applications presented in the first edition. We discuss new methods and their results. The number of research projects and publications that use social media data is constantly increasing due to continuously growing amounts of social media data and the need to automatically process them. We have added 85 new references to the more than 300 references from the first edition. Besides updating each section, we have added a new application (digital marketing) to the section on media monitoring and we have augmented the section on healthcare applications with an extended discussion of recent research on detecting signs of mental illness from social media.

Includes the proceedings of the 11th International Conference on Computational Linguistics and Intelligent Text Processing, held in Iași, Romania, in March 2010. This title covers topics such as: lexical resources, syntax and parsing, word sense disambiguation and named entity recognition, semantics and dialog, and humor and emotions.

The java New I/O (NIO) packages in J2SE 1.4 introduce many new, indispensable features previously unavailable to Java programmers. These include APIs for high-performance I/O operations, regular expression processing, and character set coding. These new libraries are a treasure trove for java developers. The NIO APIs are especially valuable where high-performance I/O is a requirement, but they can also be useful in a wide range of scenarios. The new APIs let you work directly with I/O buffers, multiplex nonblocking streams, do scattering reads and gathering writes, do channel-to-channel transfers, work with memory-mapped files, manage file locks, and much more. The new high-performance Regular Expression Library provides sophisticated, Perl-like regex-processing features such as pattern matching, search and replace, capture groups, look ahead assertions, and many others. The Charset API gives you complete control over character set encoding and decoding, which are vital for properly managing the exchange of documents on the Web, for localization, or for other purposes. You can also create and install your own custom character sets. Staying current with the latent java technology is never easy. NIO, new in Java 1.4, is quite possibly the most important new java feature since Swing. Understanding it thoroughly is essential for any serious Java developer. NIO closes the gap between java and natively compiled languages and enables java applications to achieve maximum I/O performance by effectively leveraging operating-system services in a portable way. Java NIO is a comprehensive guide to the java New I/O facilities. It lets you take full advantage of NIO features and shows you how they work, what they can do for you, and when you should use them. This book brings you up to speed on NIO and shows you how to bring your I/O-bound Java applications up to speed as well. Java NIO is an essential part of any Java professional's library.

This is the ideal introduction for students seeking to collect and analyze textual data from online sources. It covers the most critical issues that they must take into consideration at all stages of their research projects.

The new edition of an introduction to computer programming within the context of the visual arts, using the open-source programming language Processing; thoroughly updated throughout. The visual arts are rapidly changing as media moves into the web, mobile devices, and architecture. When designers and artists learn the basics of writing software, they develop a new form of literacy

that enables them to create new media for the present, and to imagine future media that are beyond the capacities of current software tools. This book introduces this new literacy by teaching computer programming within the context of the visual arts. It offers a comprehensive reference and text for Processing (www.processing.org), an open-source programming language that can be used by students, artists, designers, architects, researchers, and anyone who wants to program images, animation, and interactivity. Written by Processing's cofounders, the book offers a definitive reference for students and professionals. Tutorial chapters make up the bulk of the book; advanced professional projects from such domains as animation, performance, and installation are discussed in interviews with their creators. This second edition has been thoroughly updated. It is the first book to offer in-depth coverage of Processing 2.0 and 3.0, and all examples have been updated for the new syntax. Every chapter has been revised, and new chapters introduce new ways to work with data and geometry. New "synthesis" chapters offer discussion and worked examples of such topics as sketching with code, modularity, and algorithms. New interviews have been added that cover a wider range of projects. "Extension" chapters are now offered online so they can be updated to keep pace with technological developments in such fields as computer vision and electronics. Interviews with SUE.C, Larry Cuba, Mark Hansen, Lynn Hershman Leeson, Jürg Lehni, LettError, Golan Levin and Zachary Lieberman, Benjamin Maus, Manfred Mohr, Ash Nehru, Josh On, Bob Sabiston, Jennifer Steinkamp, Jared Tarbell, Steph Thirion, Robert Winter

This book shows how UNIX can be used effectively in the preparation of written documents, especially in the process of producing book-length documents, i.e. typesetting. As this book will demonstrate the tools available in the UNIX environment, it is also valuable to examine text processing in terms of problems and solutions: the problems faced by a writer undertaking a large writing project and the solutions offered by using the resources and power of a computer system.

This book follows a Cookbook style and is packed with intermediate and advanced level recipes. This book is for Java developers who have an interest in discovering new ways to quickly get the job done using a new language that shares many similarities with Java. The book's recipes start simple, therefore no previous Groovy experience is required to understand the code and the explanations accompanying the examples.

This book is a part of the Blue Book series "Research on the Development of Electronic Information Engineering Technology in China", which explores the cutting edge of natural language processing (NLP) studies. The research objects of natural language processing are evolved from words, phrases, and sentences to text, and research directions are from language analysis, language understanding, language generation, knowledge graphs, machine translation, to deep semantic understanding, and beyond. This is in line with the development trend of applications. And for another typical NLP application machine translation,

from text translation, to voice and image translation, now simultaneous interpretation, progress of technology makes the application of machine translation deeper and wider into diverse industries. This book is intended for researchers and industrial staffs who have been following the current situation and future trends of the natural language processing. Meanwhile, it also bears high value of reference for experts, scholars, and technical and engineering managers of different levels and different fields.

Our world is being revolutionized by data-driven methods: access to large amounts of data has generated new insights and opened exciting new opportunities in commerce, science, and computing applications. Processing the enormous quantities of data necessary for these advances requires large clusters, making distributed computing paradigms more crucial than ever. MapReduce is a programming model for expressing distributed computations on massive datasets and an execution framework for large-scale data processing on clusters of commodity servers. The programming model provides an easy-to-understand abstraction for designing scalable algorithms, while the execution framework transparently handles many system-level details, ranging from scheduling to synchronization to fault tolerance. This book focuses on MapReduce algorithm design, with an emphasis on text processing algorithms common in natural language processing, information retrieval, and machine learning. We introduce the notion of MapReduce design patterns, which represent general reusable solutions to commonly occurring problems across a variety of problem domains. This book not only intends to help the reader "think in MapReduce", but also discusses limitations of the programming model as well. This volume is a printed version of a work that appears in the Synthesis Digital Library of Engineering and Computer Science. Synthesis Lectures provide concise, original presentations of important research and development topics, published quickly, in digital and print formats. For more information visit www.morganclaypool.com

Explore various approaches to organize and extract useful text from unstructured data using Java Key Features Use deep learning and NLP techniques in Java to discover hidden insights in text Work with popular Java libraries such as CoreNLP, OpenNLP, and Mallet Explore machine translation, identifying parts of speech, and topic modeling Book Description Natural Language Processing (NLP) allows you to take any sentence and identify patterns, special names, company names, and more. The second edition of Natural Language Processing with Java teaches you how to perform language analysis with the help of Java libraries, while constantly gaining insights from the outcomes. You'll start by understanding how NLP and its various concepts work. Having got to grips with the basics, you'll explore important tools and libraries in Java for NLP, such as CoreNLP, OpenNLP, Neuroph, and Mallet. You'll then start performing NLP on different inputs and tasks, such as tokenization, model training, parts-of-speech and parsing trees. You'll learn about statistical machine translation,

summarization, dialog systems, complex searches, supervised and unsupervised NLP, and more. By the end of this book, you'll have learned more about NLP, neural networks, and various other trained models in Java for enhancing the performance of NLP applications. What you will learn Understand basic NLP tasks and how they relate to one another Discover and use the available tokenization engines Apply search techniques to find people, as well as things, within a document Construct solutions to identify parts of speech within sentences Use parsers to extract relationships between elements of a document Identify topics in a set of documents Explore topic modeling from a document Who this book is for Natural Language Processing with Java is for you if you are a data analyst, data scientist, or machine learning engineer who wants to extract information from a language using Java. Knowledge of Java programming is needed, while a basic understanding of statistics will be useful but not mandatory.

A problem-solution guide to encounter various NLP tasks utilizing Java open source libraries and cloud-based solutions Key Features Perform simple-to-complex NLP text processing tasks using modern Java libraries Extract relationships between different text complexities using a problem-solution approach Utilize cloud-based APIs to perform machine translation operations Book Description Natural Language Processing (NLP) has become one of the prime technologies for processing very large amounts of unstructured data from disparate information sources. This book includes a wide set of recipes and quick methods that solve challenges in text syntax, semantics, and speech tasks. At the beginning of the book, you'll learn important NLP techniques, such as identifying parts of speech, tagging words, and analyzing word semantics. You will learn how to perform lexical analysis and use machine learning techniques to speed up NLP operations. With independent recipes, you will explore techniques for customizing your existing NLP engines/models using Java libraries such as OpenNLP and the Stanford NLP library. You will also learn how to use NLP processing features from cloud-based sources, including Google and Amazon's AWS. You will master core tasks, such as stemming, lemmatization, part-of-speech tagging, and named entity recognition. You will also learn about sentiment analysis, semantic text similarity, language identification, machine translation, and text summarization. By the end of this book, you will be ready to become a professional NLP expert using a problem-solution approach to analyze any sort of text, sentences, or semantic words. What you will learn Explore how to use tokenizers in NLP processing Implement NLP techniques in machine learning and deep learning applications Identify sentences within the text and learn how to train specialized NER models Learn how to classify documents and perform sentiment analysis Find semantic similarities between text elements and extract text from a variety of sources Preprocess text from a variety of data sources Learn how to identify and translate languages Who this book is for This book is for data scientists, NLP engineers, and machine learning developers who

want to perform their work on linguistic applications faster with the use of popular libraries on JVM machines. This book will help you build real-world NLP applications using a recipe-based approach. Prior knowledge of Natural Language Processing basics and Java programming is expected.

This book is for experienced Java developers with NLP needs, whether academics, industrialists, or hobbyists. A basic knowledge of NLP terminology will be beneficial.

This book constitutes the proceedings of the 5th International Conference on Technologies and Innovation, CITI 2019, held in Guayaquil, Ecuador, in December 2019. The 14 full papers presented in this volume were carefully reviewed and selected from 32 submissions. They are organized in topical sections named: ICT in agronomy; knowledge-based systems and pattern recognition; internet of things and computer architecture.

Summary Taming Text, winner of the 2013 Jolt Awards for Productivity, is a hands-on, example-driven guide to working with unstructured text in the context of real-world applications. This book explores how to automatically organize text using approaches such as full-text search, proper name recognition, clustering, tagging, information extraction, and summarization. The book guides you through examples illustrating each of these topics, as well as the foundations upon which they are built. About this Book There is so much text in our lives, we are practically drowning in it. Fortunately, there are innovative tools and techniques for managing unstructured information that can throw the smart developer a much-needed lifeline. You'll find them in this book. Taming Text is a practical, example-driven guide to working with text in real applications. This book introduces you to useful techniques like full-text search, proper name recognition, clustering, tagging, information extraction, and summarization. You'll explore real use cases as you systematically absorb the foundations upon which they are built. Written in a clear and concise style, this book avoids jargon, explaining the subject in terms you can understand without a background in statistics or natural language processing. Examples are in Java, but the concepts can be applied in any language. Written for Java developers, the book requires no prior knowledge of GWT. 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. Winner of 2013 Jolt Awards: The Best Books—one of five notable books every serious programmer should read. What's Inside When to use text-taming techniques Important open-source libraries like Solr and Mahout How to build text-processing applications About the Authors Grant Ingersoll is an engineer, speaker, and trainer, a Lucene committer, and a cofounder of the Mahout machine-learning project. Thomas Morton is the primary developer of OpenNLP and Maximum Entropy. Drew Farris is a technology consultant, software developer, and contributor to Mahout, Lucene, and Solr. "Takes the mystery out of very complex processes."—From the Foreword by Liz Liddy, Dean, iSchool, Syracuse University Table of Contents Getting started taming text Foundations of taming text Searching Fuzzy string matching Identifying people, places, and things Clustering text Classification, categorization, and tagging Building an example question answering system Untamed text: exploring the next frontier

This revised and expanded new edition of an internationally successful classic presents an accessible introduction to the key methods in digital image processing for both practitioners and teachers. Emphasis is placed on practical application, presenting precise algorithmic descriptions in an unusually high level of detail, while highlighting direct connections between the mathematical foundations and concrete implementation. The text is supported by practical examples and carefully constructed chapter-ending exercises drawn from the authors' years of teaching experience, including easily adaptable Java code and completely worked out examples. Source code, test images and additional instructor materials are also provided at an

associated website. Digital Image Processing is the definitive textbook for students, researchers, and professionals in search of critical analysis and modern implementations of the most important algorithms in the field, and is also eminently suitable for self-study.

Text Processing in Java

Written in an informal, conversational, and humorous style, the second edition of Introduction to Programming Using Processing makes learning programming a fun experience. The freely-available Processing language is ideal for a first course in programming. The simple-to-access graphics and multimedia capabilities of the language let students develop eye-catching, animated programs, instead of traditional programs that print text to the console. The text takes a "classes-later" approach, focusing on basics, using objects, selection, iteration, topdown design, and arrays, before writing classes. Every example is presented in the context of the RADIS (Requirements / Analyze / Design / Implement / Support) framework, with considerable attention paid to design. Other positive habits, like good commenting practice and coding style, are emphasized as well.

Use Java to create a diverse range of Data Science applications and bring Data Science into production About This Book An overview of modern Data Science and Machine Learning libraries available in Java Coverage of a broad set of topics, going from the basics of Machine Learning to Deep Learning and Big Data frameworks. Easy-to-follow illustrations and the running example of building a search engine. Who This Book Is For This book is intended for software engineers who are comfortable with developing Java applications and are familiar with the basic concepts of data science. Additionally, it will also be useful for data scientists who do not yet know Java but want or need to learn it. If you are willing to build efficient data science applications and bring them in the enterprise environment without changing the 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 Find out how to approach different machine learning problems with Java Process unstructured information such as natural language text or images Create your own search engine Get state-of-the-art performance with XGBoost Learn how to build deep neural networks with DeepLearning4j Build applications that scale and process large amounts of data Deploy data science models to production and evaluate their performance In Detail Java is the most popular programming language, according to the TIOBE index, and it is a typical choice for running production systems in many companies, both in the startup world and among large enterprises. Not surprisingly, it is also a common choice for creating data science applications: it is fast and has a great set of data processing tools, both built-in and external. What is more, choosing Java for data science allows you to easily integrate solutions with existing software, and bring data science into production with less effort. This book will teach you how to create data science applications with Java. First, we will revise the most important things when starting a data science application, and then brush up the basics of Java and machine learning before diving into more advanced topics. We start by going over the existing libraries for data processing and libraries with machine learning algorithms. After that, we cover topics such as classification and regression, dimensionality reduction and clustering, information retrieval and natural language processing, and deep learning and big data. Finally, we finish the book by talking about the ways to deploy the model and evaluate it in production settings. Style and approach This is a practical guide where all the important concepts such as classification, regression, and dimensionality reduction are explained with the help of examples.

"In this course we will cover the essence of natural language processing (NLP) using Java. This video course will commence by walking you through basic NLP tasks including data acquisition, data cleaning, finding parts of text, and determining the end of sentences. These serve as the basis for other NLP tasks such as classifying text and determining the relationship between text elements. This will be followed by the use of tokenization techniques.

Tokenization is used for almost all NLP tasks. You will learn how text can be split to reveal information such as names, dates, and even the grammatical structure of a sentence. These types of activity can lead to insights into the relationships between text elements and embedded meaning in a document. "--Resource description page.

Examine the techniques and Java tools supporting the growing field of data science About This Book Your entry ticket to the world of data science with the stability and power of Java Explore, analyse, and visualize your data effectively using easy-to-follow examples Make your Java applications more capable using machine learning Who This Book Is For This book is for Java developers who are comfortable developing applications in Java. Those who now want to enter the world of data science or wish to build intelligent applications will find this book ideal.

Aspiring data scientists will also find this book very helpful. What You Will Learn Understand the nature and key concepts used in the field of data science Grasp how data is collected, cleaned, and processed Become comfortable with key data analysis techniques See specialized analysis techniques centered on machine learning Master the effective visualization of your data Work with the Java APIs and techniques used to perform data analysis In Detail Data science is concerned with extracting knowledge and insights from a wide variety of data sources to analyse patterns or predict future behaviour. It draws from a wide array of disciplines including statistics, computer science, mathematics, machine learning, and data mining. In this book, we cover the important data science concepts and how they are supported by Java, as well as the often statistically challenging techniques, to provide you with an understanding of their purpose and application. The book starts with an introduction of data science, followed by the basic data science tasks of data collection, data cleaning, data analysis, and data visualization. This is followed by a discussion of statistical techniques and more advanced topics including machine learning, neural networks, and deep learning. The next section examines the major categories of data analysis including text, visual, and audio data, followed by a discussion of resources that support parallel implementation. The final chapter illustrates an in-depth data science problem and provides a comprehensive, Java-based solution. Due to the nature of the topic, simple examples of techniques are presented early followed by a more detailed treatment later in the book. This permits a more natural introduction to the techniques and concepts presented in the book. Style and approach This book follows a tutorial approach, providing examples of each of the major concepts covered. With a step-by-step instructional style, this book covers various facets of data science and will get you up and running quickly.

Step-by-step guide that introduces novices to using all major features of Eclipse 3 Eclipse is an open source extensible integrated development environment (IDE) that helps Java programmers build best-of-breed integrated tools covering the whole software lifecycle-from conceptual modeling to deployment Eclipse is fast becoming the development platform of choice for the Java community Packed with code-rich, real-world examples that show programmers how to speed up the development of applications by reusing and extending existing Eclipse components Describes SWT and JFace (Eclipse's alternative to the Java AWT and Swing) and demonstrates them in practice in a JavaLayer based MP3 player Shows how Eclipse can be used as a tool platform and application framework

From lambda expressions and JavaFX 8 to new support for network programming and mobile development, Java 8 brings a wealth of changes. This cookbook helps you get up to speed right away with hundreds of hands-on recipes across a broad range of Java topics. You'll learn useful techniques for everything from debugging and data structures to GUI development and functional programming. Each recipe includes self-contained code solutions that you can freely use, along with a discussion of how and why they work. If you are

familiar with Java basics, this cookbook will bolster your knowledge of the language in general and Java 8's main APIs in particular. Recipes include: Methods for compiling, running, and debugging Manipulating, comparing, and rearranging text Regular expressions for string- and pattern-matching Handling numbers, dates, and times Structuring data with collections, arrays, and other types Object-oriented and functional programming techniques Directory and filesystem operations Working with graphics, audio, and video GUI development, including JavaFX and handlers Network programming on both client and server Database access, using JPA, Hibernate, and JDBC Processing JSON and XML for data storage Multithreading and concurrency

Specialised linguistic research needs can no longer be met by available software. This book enables the researcher to write programs for text and corpus processing, using the popular and easy to learn Java language.

In recent years, online social networking has revolutionized interpersonal communication. The newer research on language analysis in social media has been increasingly focusing on the latter's impact on our daily lives, both on a personal and a professional level. Natural language processing (NLP) is one of the most promising avenues for social media data processing. It is a scientific challenge to develop powerful methods and algorithms that extract relevant information from a large volume of data coming from multiple sources and languages in various formats or in free form. This book will discuss the challenges in analyzing social media texts in contrast with traditional documents. Research methods in information extraction, automatic categorization and clustering, automatic summarization and indexing, and statistical machine translation need to be adapted to a new kind of data. This book reviews the current research on NLP tools and methods for processing the non-traditional information from social media data that is available in large amounts, and it shows how innovative NLP approaches can integrate appropriate linguistic information in various fields such as social media monitoring, health care, and business intelligence. The book further covers the existing evaluation metrics for NLP and social media applications and the new efforts in evaluation campaigns or shared tasks on new datasets collected from social media. Such tasks are organized by the Association for Computational Linguistics (such as SemEval tasks), the National Institute of Standards and Technology via the Text REtrieval Conference (TREC) and the Text Analysis Conference (TAC), or the Conference and Labs of the Evaluation Forum (CLEF). In this third edition of the book, the authors added information about recent progress in NLP for social media applications, including more about the modern techniques provided by deep neural networks (DNNs) for modeling language and analyzing social media data. Haskell in Depth unlocks a new level of skill with this challenging language. Going beyond the basics of syntax and structure, this book opens up critical topics like advanced types, concurrency, and data processing. Summary Turn the corner from "Haskell student" to "Haskell developer." Haskell in Depth

explores the important language features and programming skills you'll need to build production-quality software using Haskell. And along the way, you'll pick up some interesting insights into why Haskell looks and works the way it does. Get ready to go deep! Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the technology Software for high-precision tasks like financial transactions, defense systems, and scientific research must be absolutely, provably correct. As a purely functional programming language, Haskell enforces a mathematically rigorous approach that can lead to concise, efficient, and bug-free code. To write such code you'll need deep understanding. You can get it from this book! About the book Haskell in Depth unlocks a new level of skill with this challenging language. Going beyond the basics of syntax and structure, this book opens up critical topics like advanced types, concurrency, and data processing. You'll discover key parts of the Haskell ecosystem and master core design patterns that will transform how you write software. What's inside Building applications, web services, and networking apps Using sophisticated libraries like lens, singletons, and servant Organizing projects with Cabal and Stack Error-handling and testing Pure parallelism for multicore processors About the reader For developers familiar with Haskell basics. About the author Vitaly Bragilevsky has been teaching Haskell and functional programming since 2008. He is a member of the GHC Steering Committee. Table of Contents PART 1 CORE HASKELL 1 Functions and types 2 Type classes 3 Developing an application: Stock quotes PART 2 INTRODUCTION TO APPLICATION DESIGN 4 Haskell development with modules, packages, and projects 5 Monads as practical functionality providers 6 Structuring programs with monad transformers PART 3 QUALITY ASSURANCE 7 Error handling and logging 8 Writing tests 9 Haskell data and code at run time 10 Benchmarking and profiling PART 4 ADVANCED HASKELL 11 Type system advances 12 Metaprogramming in Haskell 13 More about types PART 5 HASKELL TOOLKIT 14 Data-processing pipelines 15 Working with relational databases 16 Concurrency

The free book "Fundamentals of Computer Programming with C#" is a comprehensive computer programming tutorial that teaches programming, logical thinking, data structures and algorithms, problem solving and high quality code with lots of examples in C#. It starts with the first steps in programming and software development like variables, data types, conditional statements, loops and arrays and continues with other basic topics like methods, numeral systems, strings and string processing, exceptions, classes and objects. After the basics this fundamental programming book enters into more advanced programming topics like recursion, data structures (lists, trees, hash-tables and graphs), high-quality code, unit testing and refactoring, object-oriented principles (inheritance, abstraction, encapsulation and polymorphism) and their implementation the C# language. It also covers fundamental topics that each good developer should know like algorithm design, complexity of algorithms and problem solving. The

book uses C# language and Visual Studio to illustrate the programming concepts and explains some C# / .NET specific technologies like lambda expressions, extension methods and LINQ. The book is written by a team of developers lead by Svetlin Nakov who has 20+ years practical software development experience. It teaches the major programming concepts and way of thinking needed to become a good software engineer and the C# language in the meantime. It is a great start for anyone who wants to become a skillful software engineer. The books does not teach technologies like databases, mobile and web development, but shows the true way to master the basics of programming regardless of the languages, technologies and tools. It is good for beginners and intermediate developers who want to put a solid base for a successful career in the software engineering industry. The book is accompanied by free video lessons, presentation slides and mind maps, as well as hundreds of exercises and live examples. Download the free C# programming book, videos, presentations and other resources from <http://introprogramming.info>. Title: Fundamentals of Computer Programming with C# (The Bulgarian C# Programming Book) ISBN: 9789544007737 ISBN-13: 978-954-400-773-7 (9789544007737) ISBN-10: 954-400-773-3 (9544007733) Author: Svetlin Nakov & Co. Pages: 1132 Language: English Published: Sofia, 2013 Publisher: Faber Publishing, Bulgaria Web site: <http://www.introprogramming.info> License: CC-Attribution-Share-Alike Tags: free, programming, book, computer programming, programming fundamentals, ebook, book programming, C#, CSharp, C# book, tutorial, C# tutorial; programming concepts, programming fundamentals, compiler, Visual Studio, .NET, .NET Framework, data types, variables, expressions, statements, console, conditional statements, control-flow logic, loops, arrays, numeral systems, methods, strings, text processing, StringBuilder, exceptions, exception handling, stack trace, streams, files, text files, linear data structures, list, linked list, stack, queue, tree, balanced tree, graph, depth-first search, DFS, breadth-first search, BFS, dictionaries, hash tables, associative arrays, sets, algorithms, sorting algorithm, searching algorithms, recursion, combinatorial algorithms, algorithm complexity, OOP, object-oriented programming, classes, objects, constructors, fields, properties, static members, abstraction, interfaces, encapsulation, inheritance, virtual methods, polymorphism, cohesion, coupling, enumerations, generics, namespaces, UML, design patterns, extension methods, anonymous types, lambda expressions, LINQ, code quality, high-quality code, high-quality classes, high-quality methods, code formatting, self-documenting code, code refactoring, problem solving, problem solving methodology, 9789544007737, 9544007733

Provides information on data analysis from a vareity of social networking sites, including Facebook, Twitter, and LinkedIn.

If you are a Java programmer who wants to learn about the fundamental tasks underlying natural language processing, this book is for you. You will be able to identify and use NLP tasks for many common problems, and integrate them in

your applications to solve more difficult problems. Readers should be familiar/experienced with Java software development.

Scientific Data Analysis using Jython Scripting and Java presents practical approaches for data analysis using Java scripting based on Jython, a Java implementation of the Python language. The chapters essentially cover all aspects of data analysis, from arrays and histograms to clustering analysis, curve fitting, metadata and neural networks. A comprehensive coverage of data visualisation tools implemented in Java is also included. Written by the primary developer of the jHepWork data-analysis framework, the book provides a reliable and complete reference source laying the foundation for data-analysis applications using Java scripting. More than 250 code snippets (of around 10-20 lines each) written in Jython and Java, plus several real-life examples help the reader develop a genuine feeling for data analysis techniques and their programming implementation. This is the first data-analysis and data-mining book which is completely based on the Jython language, and opens doors to scripting using a fully multi-platform and multi-threaded approach. Graduate students and researchers will benefit from the information presented in this book.

Explore various approaches to organize and extract useful text from unstructured data using Java Key Features Use deep learning and NLP techniques in Java to discover hidden insights in text Work with popular Java libraries such as CoreNLP, OpenNLP, and Mallet Explore machine translation, identifying parts of speech, and topic modeling Book Description Natural Language Processing (NLP) allows you to take any sentence and identify patterns, special names, company names, and more. The second edition of Natural Language Processing with Java teaches you how to perform language analysis with the help of Java libraries, while constantly gaining insights from the outcomes. You'll start by understanding how NLP and its various concepts work. Having got to grips with the basics, you'll explore important tools and libraries in Java for NLP, such as CoreNLP, OpenNLP, Neuroph, and Mallet. You'll then start performing NLP on different inputs and tasks, such as tokenization, model training, parts-of-speech and parsing trees. You'll learn about statistical machine translation, summarization, dialog systems, complex searches, supervised and unsupervised NLP, and more. By the end of this book, you'll have learned more about NLP, neural networks, and various other trained models in Java for enhancing the performance of NLP applications. What you will learn Understand basic NLP tasks and how they relate to one another Discover and use the available tokenization engines Apply search techniques to find people, as well as things, within a document Construct solutions to identify parts of speech within sentences Use parsers to extract relationships between elements of a document Identify topics in a set of documents Explore topic modeling from a document Who this book is for Natural Language Processing with Java is for you if you are a data analyst, data scientist, or machine learning engineer who wants to extract information from a language using Java. Knowledge of Java programming is

needed, while a basic understanding of statistics will be useful but not mandatory. Downloading the example code for this book You can download the example code files for all Packt books you have purchased from your account at <http://www.PacktPub.com>. If you purchased this book elsewhere, you can visit <http://www.PacktPub.com/support> and register to have the files e-mailed directly to you.

formats using XSLT transformations. The two main text analytics architectures, GATE and UIMA, are then described and compared, with practical exercises showing how to configure and customize them. The final chapter is an introduction to text analytics, describing the main applications and functions including named entity recognition, coreference resolution and information extraction, with practical examples using both open source and commercial tools." --Book Jacket.

This book offers a highly accessible introduction to natural language processing, the field that supports a variety of language technologies, from predictive text and email filtering to automatic summarization and translation. With it, you'll learn how to write Python programs that work with large collections of unstructured text. You'll access richly annotated datasets using a comprehensive range of linguistic data structures, and you'll understand the main algorithms for analyzing the content and structure of written communication. Packed with examples and exercises, Natural Language Processing with Python will help you: Extract information from unstructured text, either to guess the topic or identify "named entities" Analyze linguistic structure in text, including parsing and semantic analysis Access popular linguistic databases, including WordNet and treebanks Integrate techniques drawn from fields as diverse as linguistics and artificial intelligence This book will help you gain practical skills in natural language processing using the Python programming language and the Natural Language Toolkit (NLTK) open source library. If you're interested in developing web applications, analyzing multilingual news sources, or documenting endangered languages -- or if you're simply curious to have a programmer's perspective on how human language works -- you'll find Natural Language Processing with Python both fascinating and immensely useful.

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.

When Lucene first hit the scene five years ago, it was nothing short of amazing. 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" feature into an indispensable part of most enterprise applications. Lucene now powers search in diverse companies including Akamai, Netflix, LinkedIn, Technorati, HotJobs, Epiphany, FedEx, Mayo Clinic, MIT, New Scientist Magazine, and many others. Some things remain the same, though. Lucene still delivers high-performance search features in a disarmingly easy-to-use API. Due to its vibrant and diverse open-source community of

developers and users, Lucene is relentlessly improving, with evolutions to APIs, significant new features such as payloads, and a huge increase (as much as 8x) in indexing speed with Lucene 2.3. And with clear writing, reusable examples, and unmatched advice on best practices, Lucene in Action, Second Edition is still the definitive guide to developing 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.

Learning Processing, Second Edition, is a friendly start-up guide to Processing, a free, open-source alternative to expensive software and daunting programming languages. Requiring no previous experience, this book is for the true programming beginner. It teaches the basic building blocks of programming needed to create cutting-edge graphics applications including interactive art, live video processing, and data visualization. Step-by-step examples, thorough explanations, hands-on exercises, and sample code, supports your learning curve. A unique lab-style manual, the book gives graphic and web designers, artists, and illustrators of all stripes a jumpstart on working with the Processing programming environment by providing instruction on the basic principles of the language, followed by careful explanations of select advanced techniques. The book has been developed with a supportive learning experience at its core. From algorithms and data mining to rendering and debugging, it teaches object-oriented programming from the ground up within the fascinating context of interactive visual media. This book is ideal for graphic designers and visual artists without programming background who want to learn programming. It will also appeal to students taking college and graduate courses in interactive media or visual computing, and for self-study. A friendly start-up guide to Processing, a free, open-source alternative to expensive software and daunting programming languages. No previous experience required—this book is for the true programming beginner! Step-by-step examples, thorough explanations, hands-on exercises, and sample code supports your learning curve.

[Copyright: 518713b17b85441b4b82487f9dfc978d](#)