

Building Search Applications Lucene Lingpipe And Gate

This book constitutes the proceedings of the 23rd International Conference on Theory and Practice of Digital Libraries, TPD 2019, held in Oslo, Norway, in September 2019. The 16 revised full papers, 12 short papers and 18 poster papers presented were carefully reviewed and selected from 75 submissions. The general theme of TPD 2019 was Connecting with Communities and so the papers attempt to facilitate establishing connections and convergences between diverse research communities such as Digital Humanities, Information Sciences and others that could benefit from ecosystems offered by digital libraries and repositories. To become especially useful to the diverse research and practitioner communities digital libraries need to consider special needs and requirements for effective data utilization, management and exploitation.

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.

The rapid advancement of semantic web technologies, along with the fact that they are at various levels of maturity, has left many practitioners confused about the current state of these technologies. Focusing on the most mature technologies, Applied Semantic Web Technologies integrates theory with case studies to illustrate the history, current state, and future direction of the semantic web. It maintains an emphasis on real-world applications and examines the technical and practical issues related to the use of semantic technologies in intelligent information management. The book starts with an introduction to the fundamentals—reviewing ontology basics, ontology languages, and research related to ontology alignment, mediation, and mapping. Next, it covers ontology engineering issues and presents a collaborative ontology engineering tool that is an extension of the Semantic MediaWiki. Unveiling a novel approach to data and knowledge engineering, the text: Introduces cutting-edge taxonomy-aware algorithms Examines semantics-based service composition in transport logistics Offers ontology alignment tools that use information visualization techniques Explains how to enrich the representation of entity semantics in an ontology Addresses challenges in tackling the content creation bottleneck Using case studies, the book provides authoritative insights and highlights valuable lessons learned by the authors—information systems veterans with decades of experience. They explain how to create social ontologies and present examples of the application of semantic technologies in building automation, logistics, ontology-driven business process intelligence, decision making, and energy efficiency in smart homes.

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.

Multilingual Natural Language Processing Applications is the first comprehensive single-source guide to building robust and accurate multilingual NLP systems. Edited by two leading experts, it integrates cutting-edge advances with practical solutions drawn from extensive field experience. Part I introduces the core concepts and theoretical foundations of modern multilingual natural language processing, presenting today's best practices for understanding word and document structure, analyzing syntax, modeling language, recognizing entailment, and detecting redundancy. Part II thoroughly addresses the practical considerations associated with building real-world applications, including information extraction, machine translation, information retrieval/search, summarization, question answering, distillation, processing pipelines, and more. This book contains important new contributions from leading researchers at IBM, Google, Microsoft, Thomson Reuters, BBN, CMU, University of Edinburgh, University of Washington, University of North Texas, and others. Coverage includes Core NLP problems, and today's best algorithms for attacking them Processing the diverse morphologies present in the world's languages

Uncovering syntactical structure, parsing semantics, using semantic role labeling, and scoring grammaticality Recognizing inferences, subjectivity, and opinion polarity Managing key algorithmic and design tradeoffs in real-world applications Extracting information via mention detection, coreference resolution, and events Building large-scale systems for machine translation, information retrieval, and summarization Answering complex questions through distillation and other advanced techniques Creating dialog systems that leverage advances in speech recognition, synthesis, and dialog management Constructing common infrastructure for multiple multilingual text processing applications This book will be invaluable for all engineers, software developers, researchers, and graduate students who want to process large quantities of text in multiple languages, in any environment: government, corporate, or academic.

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.

This book will help you: -- Learn words from a pre-defined or custom word list. A ranked list of 8500 words for the SAT and GRE exams is included. -- Create spelling and vocabulary quizzes from your word lists. Test your vocabulary with auto-generated questions to learn words and meanings. -- Build your vocabulary with a collection of word games including Hangman, guess the word, unscramble the word, find the missing word in a phrase, and select a word relationship. -- Write essays for automated essay evaluation. Create essays that will be scored on an evaluator like the E-rater from ETS. -- Practice sentence completion with dynamic quizzes customized to your skill level. -- Study grammar with sentence quizzes that test your knowledge of specific grammar rules.

This book contains a selection of papers from The 2015 International Conference on Software Process Improvement (CIMPS'15), held between the 28th and 30th of October in Mazatlán, Sinaloa, México. The CIMPS'15 is a global forum for researchers and practitioners that present and discuss the most recent innovations, trends, results, experiences and concerns in the several perspectives of Software Engineering with clear relationship but not limited to software processes, Security in Information and Communication Technology and Big Data Field. The main topics covered are: Organizational Models, Standards and Methodologies, Knowledge Management, Software Systems, Applications and Tools, Information and Communication Technologies and Processes in non-software domains (Mining, automotive, aerospace, business, health care, manufacturing, etc.) with a demonstrated relationship to software process challenges.

This volume contains papers presented at the 2nd International Afro-European Conference for Industrial Advancement -- AECIA 2015. The conference aimed at bringing together the foremost experts and excellent young researchers from Africa, Europe and the rest of the world to disseminate the latest results from various fields of engineering, information, and communication technologies. The topics, discussed at the conference, covered a broad range of domains spanning from ICT and engineering to prediction, modeling, and analysis of complex systems. The 2015 edition of AECIA featured a distinguished special track on prediction, modeling and analysis of complex systems -- Nostradamus, and special sessions on Advances in Image Processing and Colorization and Data Processing, Protocols, and Applications in Wireless Sensor Networks.

This book is written for anyone who is interested in how a field of research evolves and the fundamental role of understanding uncertainties involved in different levels of analysis, ranging from macroscopic views to meso- and microscopic ones. We introduce a series of computational and visual analytic techniques, from research areas such as text mining, deep learning, information visualization and science mapping, such that readers can apply these tools to the study of a subject matter of their choice. In addition, we set the diverse set of methods in an integrative context, that draws upon insights from philosophical, sociological, and evolutionary theories of what drives the advances of science, such that the readers of the book can guide their own research with their enriched theoretical foundations. Scientific knowledge is complex. A subject matter is typically built on its own set of concepts, theories, methodologies and findings, discovered by generations of researchers and practitioners. Scientific knowledge, as known to the scientific community as a whole, experiences constant changes. Some changes are long-lasting, whereas others may be short lived. How can we keep abreast of the state of the art as science advances? How can we effectively and precisely convey the status of the current science to the general public as well as scientists across different disciplines? The study of scientific knowledge in general has been overwhelmingly focused on scientific knowledge per se. In contrast, the status of scientific knowledge at various levels of granularity has been largely overlooked. This book aims to highlight the role of uncertainties, in developing a better

understanding of the status of scientific knowledge at a particular time, and how its status evolves over the course of the development of research. Furthermore, we demonstrate how the knowledge of the types of uncertainties associated with scientific claims serves as an integral and critical part of our domain expertise.

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.

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

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.

THIRD EDITION: Did you learn PowerPoint in 30 minutes? Join the crowd...most people get no more than a half-hour of training time with PowerPoint before they are tasked with making what is likely to be a first impression of themselves or their company. This book is for earnest presenters and presentation designers who want to escape the perils that entrap so many who turn to PowerPoint for their presentations.

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.

A guide for geographic analysts, modelers, software engineers, and GIS professionals, this book discusses agent-based modeling, dynamic feedback and simulation modeling, as well as links between models and GIS software. This collection also presents a state-of-the-art understanding of applications based on environmental, atmospheric, hydrological, urban, social, health, and economic models.

GATE is a free open-source infrastructure for developing and deploying software components that process human language. It is more than 15 years old and is in active use for all types of computational tasks involving language (frequently called natural language processing, text analytics, or text mining). GATE excels at text analysis of all shapes and sizes. From large corporations to small startups, from multi-million research consortia to undergraduate projects, our user community is the largest and most diverse of any system of this type, and is active world-wide. This book contains a highly accessible introduction to GATE Version 6 and is the first port of call for all GATE-related questions. It includes a guide to using GATE Developer and GATE Embedded, and chapters on all major areas of functionality, such as processing multiple languages and large collections of unstructured text. It also includes complete plugin documentation (e.g. named entity recognition, parsing, semantic analysis, , as well as details on other members of the GATE Family: GATECloud.net, Teamware, and Mimir. To join the GATE community visit <http://gate.ac.uk/>.

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.

Biomedical Natural Language Processing is a comprehensive tour through the classic and current work in the field. It discusses all subjects from both a rule-based and a machine learning approach, and also describes each subject from the perspective of both biological science and clinical medicine. The intended audience is readers who already have a background in natural language processing, but a clear introduction makes it accessible to readers from the fields of bioinformatics and computational biology, as well. The book is suitable as a reference, as well as a text for advanced courses in biomedical natural language processing and text mining.

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.

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.

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.

With Google Web Toolkit, Java developers can build sophisticated Rich Internet Applications (RIAs) and complete Web sites using the powerful IDEs and tools they already use. Now, with GWT 2, Google Web Toolkit has become even more useful. Essential GWT shows how to use this latest version of GWT to create production solutions that combine superior style, performance, and interactivity with exceptional quality and maintainability. Federico Kereki quickly reviews the basics and then introduces intermediate and advanced GWT skills, covering issues ranging from organizing projects to compiling and deploying final code. Throughout, he focuses on best-practice methodologies and design patterns. For example, you'll learn how to use the MVP (model-view-presenter) pattern to improve application design and support automated testing for agile development. Kereki illuminates each concept with realistic code examples that help developers jump-start their projects and get great results more quickly. Working with the latest versions of open source tools such as Eclipse, Subversion, Apache, Tomcat, and MySQL, he demonstrates exactly how GWT fits into real Web development environments. Coverage includes Using the Google Plugin for Eclipse and the GWT Shell Script Detecting and working with browsers—and solving the problems they cause Building better user interfaces with the MVP pattern Using APIs for visualization, mapping, weather data, and more Internationalizing and localizing GWT code Securing GWT applications with cryptography, hashing, and encryption Testing with JUnit, Emma, GWTTestCase, Selenium, and Mock Objects Deploying client-only and client-plus-server GWT applications

Summary Kafka Streams in Action teaches you everything you need to know to implement stream processing on data flowing into your Kafka platform, allowing you to focus on getting more from your data without sacrificing time or effort. Foreword by Neha Narkhede, Cocreator of Apache Kafka Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Not all stream-based applications require a dedicated processing cluster. The lightweight Kafka Streams library provides exactly the power and simplicity you need for message handling in microservices and real-time event processing. With the Kafka Streams API, you filter and transform data streams with just Kafka and your application. About the Book Kafka Streams in Action teaches you to implement stream processing within the Kafka platform. In this easy-to-follow book, you'll explore real-world examples to collect, transform, and aggregate data, work with multiple processors, and handle real-time events. You'll even dive into streaming SQL with KSQL! Practical to the very end, it finishes with testing and operational aspects, such as monitoring and debugging. What's inside Using the KStreams API Filtering, transforming, and splitting data Working with the Processor API Integrating with external systems About the Reader Assumes some experience with distributed systems. No

knowledge of Kafka or streaming applications required. About the Author Bill Bejeck is a Kafka Streams contributor and Confluent engineer with over 15 years of software development experience. Table of Contents PART 1 - GETTING STARTED WITH KAFKA STREAMS Welcome to Kafka Streams Kafka quicklyPART 2 - KAFKA STREAMS DEVELOPMENT Developing Kafka Streams Streams and state The KTable API The Processor APIPART 3 - ADMINISTERING KAFKA STREAMS Monitoring and performance Testing a Kafka Streams applicationPART 4 - ADVANCED CONCEPTS WITH KAFKA STREAMS Advanced applications with Kafka StreamsAPPENDIXES Appendix A - Additional configuration information Appendix B - Exactly once semantics

For over seven years, computer networking and security instructor and consultant, Dru Lavigne, meticulously documented her learning experiences with FreeBSD administration and open source software usage in a series of over 110 articles. Many readers praised and recommended the author's informative tutorials. The Best of FreeBSD Basics book contains most of these articles - many updated from FreeBSD 4 and 5 to reflect the usage on FreeBSD 6 and 7. The Best of FreeBSD Basics provides practical advice for completing common tasks on FreeBSD and is a great way to get to know FreeBSD - and Unix in general. Darwin, DragonFly, Linux, Mac OS X, NetBSD, and OpenBSD fans will also find a lot of the book invaluable and useful. Covering a huge range of FreeBSD and open source topics, The Best of FreeBSD Basics includes step-by-step directions, things to watch out for, and hints for success. A sampling of the book's topics include installing an X11 server and setting up an desktop environment, comparing common tasks with Linux, playing audio and video files, user administration, system startup, finding and using documentation, managing backups, networking basics, IPsec, setting up several servers, filtering spam, improving security, enabling firewalls, and a lot more.

How do you create a mission-critical site that provides exceptional performance while remaining flexible, adaptable, and reliable 24/7? Written by the manager of a UI group at Yahoo!, Developing Large Web Applications offers practical steps for building rock-solid applications that remain effective even as you add features, functions, and users. You'll learn how to develop large web applications with the extreme precision required for other types of software. Avoid common coding and maintenance headaches as small websites add more pages, more code, and more programmers Get comprehensive solutions for refining HTML, CSS, JavaScript, PHP, and Ajax for large-scale web applications Make changes in one place that ripple through all affected page elements Embrace the virtues of modularity, encapsulation, abstraction, and loosely coupled components Use tried-and-true techniques for managing data exchange, including working with forms and cookies Learn often-overlooked best practices in code management and software engineering Prepare your code to make performance enhancements and testing easier

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, 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.

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.

This book constitutes the refereed proceedings of the 9th International Conference 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.

The ways in which humans communicate with one another is constantly evolving. Technology plays a large role in this evolution via new methods and avenues of social and business interaction. Optimizing Human-Computer Interaction With Emerging Technologies is a primary reference source featuring the latest scholarly perspectives on technological breakthroughs in user operation and the processes of communication in the digital era. Including a number of topics such as health information technology, multimedia, and social media, this publication is

ideally designed for professionals, technology developers, and researchers seeking current research on technology's role in communication.

Building Search Applications with Lucene and Lingpipe Building Search Applications Lucene, LingPipe, and GateLulu.com

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.

This book constitutes the refereed proceedings of the Second International Conference on Electronic Government and the Information Systems Perspective, EGOVIS 2011, held in Toulouse, France, in August/September 2011. The 30 revised full papers presented were carefully reviewed and selected from numerous submissions. Among the topics addressed are aspects of security, reliability, privacy and anonymity of e-government systems, knowledge processing, service-oriented computing, and case studies of e-government systems in several countries.

Summary Relevant Search demystifies relevance work. Using Elasticsearch, it teaches you how to return engaging search results to your users, helping you understand and leverage the internals of Lucene-based search engines. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Users are accustomed to and expect instant, relevant search results. To achieve this, you must master the search engine. Yet for many developers, relevance ranking is mysterious or confusing. About the Book Relevant Search demystifies the subject and shows you that a search engine is a programmable relevance framework. You'll learn how to apply Elasticsearch or Solr to your business's unique ranking problems. The book demonstrates how to program relevance and how to incorporate secondary data sources, taxonomies, text analytics, and personalization. In practice, a relevance framework requires softer skills as well, such as collaborating with stakeholders to discover the right relevance requirements for your business. By the end, you'll be able to achieve a virtuous cycle of provable, measurable relevance improvements over a search product's lifetime. What's Inside Techniques for debugging relevance? Applying search engine features to real problems? Using the user interface to guide searchers? A systematic approach to relevance? A business culture focused on improving search About the Reader For developers trying to build smarter search with Elasticsearch or Solr. About the Authors Doug Turnbull is lead relevance consultant at OpenSource Connections, where he frequently speaks and blogs. John Berryman is a data engineer at Eventbrite, where he specializes in recommendations and search. Foreword author, Trey Grainger, is a director of engineering at CareerBuilder and author of Solr in Action. Table of Contents The search relevance problem Search under the hood Debugging your first relevance problem Taming tokens Basic multifield search Term-centric search Shaping the relevance function Providing relevance feedback Designing a relevance-focused search application The relevance-centered enterprise Semantic and personalized search

The Handbook of Natural Language Processing, Second Edition presents practical tools and techniques for implementing natural language processing in computer systems. Along with removing outdated material, this edition updates every chapter and expands the content to include emerging areas, such as sentiment analysis. New to the Second Edition Greater

This volume constitutes the proceedings of the 9th International Conference on Simulated Evolution and Learning, SEAL 2012, held in Hanoi, Vietnam, in December 2012. The 50 full papers presented were carefully reviewed and selected from 91 submissions. The papers are organized in topical sections on evolutionary algorithms, theoretical developments, swarm intelligence, data mining, learning methodologies, and real-world applications.

The work of black writers, editors, publishers, and librarians is deeply embedded in the history of American print culture, from slave narratives to digital databases. While the printed word can seem democratizing, it remains that the infrastructures of print and digital culture can be as limiting as they are enabling. Contributors to this volume explore the relationship between expression and such frameworks, analyzing how different mediums, library catalogs, and search engines shape the production and reception of written and visual culture. Topics include antebellum literature, the Harlem Renaissance, the Black Arts Movement; "post-Black" art, the role of black librarians, and how present-day technologies aid or hinder the discoverability of work by African Americans. Against a Sharp White Background covers elements of production, circulation, and reception of African American writing across a range of genres and contexts. This collection challenges mainstream book history and print culture to understand that race and racialization are inseparable from the study of texts and their technologies.

The field of artificial intelligence (AI) and the law is on the cusp of a revolution that began with text analytic programs like IBM's Watson and Debater and the open-source information management architectures on which they are based. Today, new legal applications are beginning to appear and this book - designed to explain computational processes to non-programmers - describes how they will change the practice of law, specifically by connecting computational models of legal reasoning directly with legal text, generating arguments for and against particular outcomes, predicting outcomes and explaining these predictions with reasons that legal professionals will be able to evaluate for themselves. These legal applications will support conceptual legal information retrieval and allow cognitive computing, enabling a collaboration between humans and computers in which each does what it can do best. Anyone interested in how AI is changing the practice of law should read this illuminating work.

[Copyright: 83bd04b40f92e4a80caba26d23ea8710](https://www.lulu.com/product/summary-relevant-search-demystifies-relevance-work/83bd04b40f92e4a80caba26d23ea8710)