

# Apache Tika Tutorial Understanding Of Apache Tika Library The File Format Content Metadata Extraction Learn Apache Tika Fast

YEAH, REVIEWING A BOOK **APACHE TIKA TUTORIAL UNDERSTANDING OF APACHE TIKA LIBRARY THE FILE FORMAT CONTENT METADATA EXTRACTION LEARN APACHE TIKA FAST** COULD GROW YOUR NEAR ASSOCIATES LISTINGS. THIS IS JUST ONE OF THE SOLUTIONS FOR YOU TO BE SUCCESSFUL. AS UNDERSTOOD, DEED DOES NOT RECOMMEND THAT YOU HAVE EXTRAORDINARY POINTS.

COMPREHENDING AS SKILLFULLY AS PACT EVEN MORE THAN NEW WILL PRESENT EACH SUCCESS. BORDERING TO, THE PRONOUNCEMENT AS CAPABLY AS INSIGHT OF THIS **APACHE TIKA TUTORIAL UNDERSTANDING OF APACHE TIKA LIBRARY THE FILE FORMAT CONTENT METADATA EXTRACTION LEARN APACHE TIKA FAST** CAN BE TAKEN AS SKILLFULLY AS PICKED TO ACT.

**INFORMATION RETRIEVAL** - DAVID A. GROSSMAN 2012-11-12

INTERESTED IN HOW AN EFFICIENT SEARCH ENGINE WORKS? WANT TO KNOW WHAT ALGORITHMS ARE USED TO RANK RESULTING DOCUMENTS IN RESPONSE TO USER REQUESTS? THE AUTHORS ANSWER THESE AND OTHER KEY INFORMATION RETRIEVAL DESIGN AND IMPLEMENTATION QUESTIONS. THIS BOOK IS NOT YET ANOTHER HIGH LEVEL TEXT. INSTEAD, ALGORITHMS ARE THOROUGHLY DESCRIBED, MAKING THIS BOOK IDEALLY SUITED FOR BOTH COMPUTER SCIENCE STUDENTS AND PRACTITIONERS WHO WORK ON SEARCH-RELATED APPLICATIONS. AS STATED IN THE FOREWORD, THIS BOOK PROVIDES A CURRENT, BROAD, AND DETAILED OVERVIEW OF THE FIELD AND IS THE ONLY ONE THAT DOES SO. EXAMPLES ARE USED THROUGHOUT TO ILLUSTRATE THE ALGORITHMS. THE AUTHORS EXPLAIN HOW A QUERY IS RANKED AGAINST A DOCUMENT COLLECTION USING EITHER A SINGLE OR A COMBINATION OF RETRIEVAL STRATEGIES, AND HOW AN ASSORTMENT OF UTILITIES ARE INTEGRATED INTO THE QUERY PROCESSING SCHEME TO IMPROVE THESE RANKINGS. METHODS FOR BUILDING AND COMPRESSING TEXT INDEXES, QUERYING AND RETRIEVING DOCUMENTS IN MULTIPLE LANGUAGES, AND USING PARALLEL OR DISTRIBUTED PROCESSING TO EXPEDITE THE SEARCH ARE LIKewise DESCRIBED. THIS EDITION IS A MAJOR EXPANSION OF THE ONE PUBLISHED IN 1998. BESIDES UPDATING THE ENTIRE BOOK WITH CURRENT TECHNIQUES, IT INCLUDES NEW SECTIONS ON LANGUAGE MODELS, CROSS-LANGUAGE INFORMATION RETRIEVAL, PEER-TO-PEER PROCESSING, XML SEARCH, MEDIATORS, AND DUPLICATE DOCUMENT DETECTION.

**TIKA IN ACTION** - JUKKA L. ZITTING 2011-11-30

SUMMARY TIKA IN ACTION IS A HANDS-ON GUIDE TO CONTENT MINING WITH APACHE TIKA. THE BOOK'S MANY EXAMPLES AND CASE STUDIES OFFER REAL-WORLD EXPERIENCE FROM DOMAINS RANGING FROM SEARCH ENGINES TO DIGITAL ASSET MANAGEMENT AND SCIENTIFIC DATA PROCESSING. ABOUT THE TECHNOLOGY TIKA IS AN APACHE TOOLKIT THAT HAS BUILT INTO IT EVERYTHING YOU AND YOUR APP NEED TO KNOW ABOUT FILE FORMATS. USING TIKA, YOUR APPLICATIONS CAN DISCOVER AND EXTRACT CONTENT FROM DIGITAL DOCUMENTS IN ALMOST ANY FORMAT, INCLUDING EXOTIC ONES. ABOUT THIS BOOK TIKA IN ACTION IS THE ULTIMATE GUIDE TO CONTENT MINING USING APACHE TIKA. YOU'LL LEARN HOW TO PULL USABLE INFORMATION FROM OTHERWISE INACCESSIBLE SOURCES, INCLUDING INTERNET MEDIA AND FILE ARCHIVES. THIS EXAMPLE-RICH BOOK TEACHES YOU TO BUILD AND EXTEND APPLICATIONS BASED ON REAL-WORLD EXPERIENCE WITH SEARCH ENGINES, DIGITAL ASSET MANAGEMENT, AND SCIENTIFIC DATA PROCESSING. IN ADDITION TO ARCHITECTURAL OVERVIEWS, YOU'LL FIND DETAILED CHAPTERS ON FEATURES LIKE METADATA EXTRACTION, AUTOMATIC LANGUAGE DETECTION, AND CUSTOM PARSER DEVELOPMENT. THIS BOOK IS WRITTEN FOR DEVELOPERS WHO ARE NEW TO BOTH SCALA AND LIFT AND COVERS JUST ENOUGH SCALA TO GET YOU STARTED. PURCHASE OF THE PRINT BOOK COMES WITH AN OFFER OF A FREE PDF, ePUB, AND KINDLE eBook FROM MANNING. ALSO AVAILABLE IS ALL CODE FROM THE BOOK. WHAT'S INSIDE CRACK MS WORD, PDF, HTML, AND ZIP INTEGRATE WITH SEARCH ENGINES, CMS, AND OTHER DATA SOURCES LEARN THROUGH EXPERIMENTATION MANY EXAMPLES THIS BOOK REQUIRES NO PREVIOUS KNOWLEDGE OF TIKA OR TEXT MINING TECHNIQUES. IT ASSUMES A WORKING KNOWLEDGE OF JAVA.

===== TABLE OF CONTENTS PART 1 GETTING STARTED THE CASE FOR THE DIGITAL BABEL FISH GETTING STARTED WITH TIKA THE INFORMATION LANDSCAPE PART 2 TIKA IN DETAIL DOCUMENT TYPE DETECTION CONTENT EXTRACTION UNDERSTANDING METADATA LANGUAGE DETECTION WHAT'S IN A FILE? PART 3 INTEGRATION AND ADVANCED USE THE BIG PICTURE TIKA AND THE LUCENE SEARCH STACK EXTENDING TIKA PART 4 CASE STUDIES POWERING NASA SCIENCE DATA SYSTEMS CONTENT MANAGEMENT WITH APACHE JACKRABBIT CURATING CANCER RESEARCH DATA WITH TIKA THE CLASSIC SEARCH ENGINE EXAMPLE

**ALFRESCO DEVELOPER GUIDE** - JEFF POTTS 2008-10-31

CUSTOMIZING ALFRESCO WITH ACTIONS, WEB SCRIPTS, WEB FORMS, WORKFLOWS, AND MORE

**DATA LAKES** - ANNE LAURENT 2020-04-09

THE CONCEPT OF A DATA LAKE IS LESS THAN 10 YEARS OLD, BUT THEY ARE ALREADY HUGELY IMPLEMENTED WITHIN LARGE COMPANIES. THEIR GOAL IS TO EFFICIENTLY DEAL WITH EVER-GROWING VOLUMES OF HETEROGENEOUS DATA, WHILE ALSO FACING VARIOUS SOPHISTICATED USER NEEDS. HOWEVER, DEFINING AND BUILDING A DATA LAKE IS STILL A CHALLENGE, AS NO CONSENSUS HAS BEEN REACHED SO FAR. DATA LAKES PRESENTS RECENT OUTCOMES AND TRENDS IN THE FIELD OF DATA REPOSITORIES. THE MAIN TOPICS DISCUSSED ARE THE DATA-DRIVEN ARCHITECTURE OF A DATA LAKE; THE MANAGEMENT OF METADATA – SUPPLYING KEY INFORMATION ABOUT THE STORED DATA, MASTER DATA AND REFERENCE DATA; THE ROLES OF LINKED DATA AND FOG COMPUTING IN A DATA LAKE ECOSYSTEM; AND HOW GRAVITY PRINCIPLES APPLY IN THE CONTEXT OF DATA LAKES. A VARIETY OF CASE STUDIES ARE ALSO PRESENTED, THUS PROVIDING THE READER WITH PRACTICAL EXAMPLES OF DATA LAKE MANAGEMENT.

**CONCEPTS AND METHODS FOR A LIBRARIAN OF THE WEB** - MARIO KUBEK 2019-06-20

THE WORLD WIDE WEB CAN BE CONSIDERED A HUGE LIBRARY THAT IN CONSEQUENCE NEEDS A CAPABLE LIBRARIAN RESPONSIBLE FOR THE CLASSIFICATION AND RETRIEVAL OF DOCUMENTS AS WELL AS THE MEDIATION BETWEEN LIBRARY RESOURCES AND USERS. BASED ON THIS IDEA, THE CONCEPT OF THE “LIBRARIAN OF THE WEB” IS INTRODUCED WHICH COMPRISES NOVEL, LIBRARIAN-INSPIRED METHODS AND TECHNICAL SOLUTIONS TO DECENTRALLY SEARCH FOR TEXT DOCUMENTS IN THE WEB USING PEER-TO-PEER TECHNOLOGY. THE CONCEPT'S IMPLEMENTATION IN THE FORM OF AN INTERACTIVE PEER-TO-PEER CLIENT, CALLED “WEBENGINE”, IS ELABORATED ON IN DETAIL. THIS SOFTWARE EXTENDS AND INTERCONNECTS COMMON WEB SERVERS CREATING A FULLY INTEGRATED, DECENTRALISED AND SELF-ORGANISING WEB SEARCH SYSTEM ON TOP OF THE EXISTING WEB STRUCTURE. THUS, THE WEB IS TURNED INTO ITS OWN POWERFUL SEARCH ENGINE WITHOUT THE NEED FOR ANY CENTRAL AUTHORITY. THIS BOOK IS INTENDED FOR RESEARCHERS AND PRACTITIONERS HAVING A SOLID BACKGROUND IN THE FIELDS OF INFORMATION RETRIEVAL AND WEB MINING.

**APACHE SOLR 4 COOKBOOK** - RAFAŁ KUJĄ 2013-01-01

OVER 100 PRACTICAL RECIPES TO MAKE APACHE SOLR FASTER, MORE RELIABLE AND RETURN BETTER RESULTS.

**NATURAL LANGUAGE ANNOTATION FOR MACHINE LEARNING** - JAMES PUSTEJOVSKY 2012-10-25

INCLUDES BIBLIOGRAPHICAL REFERENCES (P. 305-315) AND INDEX.

**CMIS AND APACHE CHEMISTRY IN ACTION** - JAY BROWN 2013-07-25

SUMMARY CMIS AND APACHE CHEMISTRY IN ACTION IS A COMPREHENSIVE GUIDE TO THE CMIS STANDARD AND RELATED ECM CONCEPTS, WRITTEN BY THE AUTHORS OF THE STANDARD. IN IT, YOU'LL TACKLE HANDS-ON EXAMPLES FOR BUILDING APPLICATIONS ON CMIS REPOSITORIES FROM BOTH THE CLIENT AND THE SERVER SIDES. YOU'LL LEARN HOW TO CREATE NEW CONTENT-CENTRIC APPLICATIONS THAT INSTALL AND RUN IN ANY CMIS-COMPLIANT REPOSITORY. ABOUT THE TECHNOLOGY CONTENT MANAGEMENT INTEROPERABILITY SERVICES (CMIS) IS AN OASIS STANDARD FOR ACCESSING CONTENT MANAGEMENT SYSTEMS. IT SPECIFIES A VENDOR- AND LANGUAGE-NEUTRAL WAY TO INTERACT WITH ANY COMPLIANT CONTENT REPOSITORY. APACHE CHEMISTRY PROVIDES COMPLETE REFERENCE IMPLEMENTATIONS OF THE CMIS STANDARD WITH ROBUST APIS FOR DEVELOPERS WRITING TOOLS, APPLICATIONS, AND SERVERS. ABOUT THIS BOOK CMIS AND APACHE CHEMISTRY IN ACTION IS A COMPREHENSIVE GUIDE TO THE CMIS STANDARD AND RELATED ECM CONCEPTS. IN IT, YOU'LL FIND CLEAR TEACHING AND INSTANTLY USEFUL EXAMPLES FOR BUILDING CONTENT-CENTRIC CLIENT AND SERVER-SIDE APPLICATIONS THAT RUN AGAINST ANY CMIS-COMPLIANT REPOSITORY. IN FACT, USING THE CMIS WORKBENCH AND THE INMEMORY REPOSITORY FROM APACHE CHEMISTRY, YOU'LL HAVE RUNNING CODE TALKING TO A REAL CMIS SERVER BY THE END OF CHAPTER 1. THIS BOOK REQUIRES SOME FAMILIARITY WITH CONTENT MANAGEMENT SYSTEMS AND A STANDARD PROGRAMMING LANGUAGE LIKE JAVA OR C#. NO EXPOSURE TO CMIS OR APACHE CHEMISTRY IS ASSUMED. PURCHASE OF THE PRINT BOOK INCLUDES A FREE eBook IN PDF, KINDLE, AND ePUB FORMATS FROM MANNING PUBLICATIONS. WHAT'S INSIDE THE ONLY CMIS BOOK ENDORSED BY OASIS COMPLETE COVERAGE OF THE CMIS 1.0 AND 1.1 SPECIFICATIONS COOKBOOK-STYLE TUTORIALS AND REAL-WORLD EXAMPLES ABOUT THE AUTHORS FLORIAN MÜLLER, JAY BROWN, AND JEFF POTTS ARE AMONG THE ORIGINAL AUTHORS, CONTRIBUTORS, AND LEADERS OF APACHE CHEMISTRY AND THE OASIS CMIS SPECIFICATION. THEY CONTINUE TO SHAPE CMIS IMPLEMENTATIONS AT ALFRESCO, IBM, AND SAP. TABLE OF CONTENTS PART 1 UNDERSTANDING CMIS INTRODUCING CMIS EXPLORING THE CMIS DOMAIN MODEL CREATING, UPDATING, AND DELETING OBJECTS WITH CMIS CMIS METADATA: TYPES AND PROPERTIES QUERY PART 2 HANDS-ON CMIS CLIENT DEVELOPMENT MEET YOUR NEW PROJECT: THE BLEND THE BLEND: READ AND QUERY FUNCTIONALITY THE BLEND: CREATE, UPDATE, AND DELETE FUNCTIONALITY USING OTHER CLIENT LIBRARIES BUILDING MOBILE APPS WITH CMIS PART 3 ADVANCED TOPICS CMIS BINDINGS SECURITY AND CONTROL PERFORMANCE BUILDING A CMIS SERVER

**SOCIAL BIG DATA ANALYTICS** - BILAL ABU-SALIH 2021-03-10

THIS BOOK FOCUSES ON DATA AND HOW MODERN BUSINESS FIRMS USE SOCIAL DATA, SPECIFICALLY ONLINE SOCIAL NETWORKS (OSNs) INCORPORATED AS PART OF THE INFRASTRUCTURE FOR A NUMBER OF EMERGING APPLICATIONS SUCH AS PERSONALIZED RECOMMENDATION SYSTEMS, OPINION ANALYSIS, EXPERTISE RETRIEVAL, AND COMPUTATIONAL ADVERTISING. THIS BOOK IDENTIFIES HOW IN SUCH APPLICATIONS, SOCIAL DATA OFFERS A PLETHORA OF BENEFITS TO ENHANCE THE DECISION MAKING PROCESS. THIS BOOK HIGHLIGHTS THAT BUSINESS INTELLIGENCE APPLICATIONS ARE MORE FOCUSED ON STRUCTURED DATA; HOWEVER, IN ORDER TO UNDERSTAND AND ANALYSE THE SOCIAL BIG DATA, THERE IS A NEED TO AGGREGATE DATA FROM VARIOUS SOURCES AND TO PRESENT IT IN A PLAUSIBLE FORMAT. BIG SOCIAL DATA (BSD) EXHIBIT ALL THE TYPICAL PROPERTIES OF BIG DATA: WIDE PHYSICAL DISTRIBUTION, DIVERSITY OF FORMATS, NON-STANDARD DATA MODELS, INDEPENDENTLY-MANAGED AND HETEROGENEOUS SEMANTICS BUT EVEN FURTHER VALUABLE WITH MARKETING OPPORTUNITIES. THE BOOK PROVIDES A REVIEW OF THE CURRENT STATE-OF-THE-ART APPROACHES FOR BIG SOCIAL

DATA ANALYTICS AS WELL AS TO PRESENT DISSIMILAR METHODS TO INFER VALUE FROM SOCIAL DATA. THE BOOK FURTHER EXAMINES SEVERAL AREAS OF RESEARCH THAT BENEFITS FROM THE PROPAGATION OF THE SOCIAL DATA. IN PARTICULAR, THE BOOK PRESENTS VARIOUS TECHNICAL APPROACHES THAT PRODUCE DATA ANALYTICS CAPABLE OF HANDLING BIG DATA FEATURES AND EFFECTIVE IN FILTERING OUT UNSOLICITED DATA AND INFERRING A VALUE. THESE APPROACHES COMPRISE ADVANCED TECHNICAL SOLUTIONS ABLE TO CAPTURE HUGE AMOUNTS OF GENERATED DATA, SCRUTINISE THE COLLECTED DATA TO ELIMINATE UNWANTED DATA, MEASURE THE QUALITY OF THE INFERRED DATA, AND TRANSFORM THE AMENDED DATA FOR FURTHER DATA ANALYSIS. FURTHERMORE, THE BOOK PRESENTS SOLUTIONS TO DERIVE KNOWLEDGE AND SENTIMENTS FROM BSD AND TO PROVIDE SOCIAL DATA CLASSIFICATION AND PREDICTION. THE APPROACHES IN THIS BOOK ALSO INCORPORATE SEVERAL TECHNOLOGIES SUCH AS SEMANTIC DISCOVERY, SENTIMENT ANALYSIS, AFFECTIVE COMPUTING AND MACHINE LEARNING. THIS BOOK HAS ADDITIONAL SPECIAL FEATURE ENRICHED WITH NUMEROUS ILLUSTRATIONS SUCH AS TABLES, GRAPHS AND CHARTS INCORPORATING ADVANCED VISUALISATION TOOLS IN ACCESSIBLE AN ATTRACTIVE DISPLAY.

**FOUNDATIONS FOR ANALYTICS WITH PYTHON** - CLINTON W. BROWNLEY 2016-08-16

IF YOU'RE LIKE MANY OF EXCEL'S 750 MILLION USERS, YOU WANT TO DO MORE WITH YOUR DATA—LIKE REPEATING SIMILAR ANALYSES OVER HUNDREDS OF FILES, OR COMBINING DATA IN MANY FILES FOR ANALYSIS AT ONE TIME. THIS PRACTICAL GUIDE SHOWS AMBITIOUS NON-PROGRAMMERS HOW TO AUTOMATE AND SCALE THE PROCESSING AND ANALYSIS OF DATA IN DIFFERENT FORMATS—BY USING PYTHON. AFTER AUTHOR CLINTON BROWNLEY TAKES YOU THROUGH PYTHON BASICS, YOU'LL BE ABLE TO WRITE SIMPLE SCRIPTS FOR PROCESSING DATA IN SPREADSHEETS AS WELL AS DATABASES. YOU'LL ALSO LEARN HOW TO USE SEVERAL PYTHON MODULES FOR PARSING FILES, GROUPING DATA, AND PRODUCING STATISTICS. NO PROGRAMMING EXPERIENCE IS NECESSARY. CREATE AND RUN YOUR OWN PYTHON SCRIPTS BY LEARNING BASIC SYNTAX USE PYTHON'S CSV MODULE TO READ AND PARSE CSV FILES READ MULTIPLE EXCEL WORKSHEETS AND WORKBOOKS WITH THE XLRD MODULE PERFORM DATABASE OPERATIONS IN MYSQL OR WITH THE MYSQLCLIENT MODULE CREATE PYTHON APPLICATIONS TO FIND SPECIFIC RECORDS, GROUP DATA, AND PARSE TEXT FILES BUILD STATISTICAL GRAPHS AND PLOTS WITH MATPLOTLIB, PANDAS, GGLOT, AND SEABORN PRODUCE SUMMARY STATISTICS, AND ESTIMATE REGRESSION AND CLASSIFICATION MODELS SCHEDULE YOUR SCRIPTS TO RUN AUTOMATICALLY IN BOTH WINDOWS AND MAC ENVIRONMENTS

**LEGAL KNOWLEDGE AND INFORMATION SYSTEMS** - A. WYNER 2017-12-20

LIKE EVERY OTHER WALK OF MODERN LIFE, THE LAW HAS EMBRACED DIGITAL TECHNOLOGY, AND IS INCREASINGLY RELIANT ON INFORMATION SYSTEMS FOR ITS EFFICIENT FUNCTIONING. THIS BOOK PRESENTS PAPERS FROM THE 30TH INTERNATIONAL CONFERENCE ON LEGAL KNOWLEDGE AND INFORMATION SYSTEMS (JURIX 2017), HELD IN LUXEMBOURG CITY, LUXEMBOURG, IN DECEMBER 2017. IN THE THREE DECADES SINCE THEY BEGAN, THE JURIX CONFERENCES HAVE BEEN HELD UNDER THE AUSPICES OF THE DUTCH FOUNDATION FOR LEGAL KNOWLEDGE BASED SYSTEMS, AND HAVE BECOME A FULLY EUROPEAN CONFERENCE SERIES WHICH ADDRESSES FAMILIAR TOPICS AND EXTENDS KNOWN TECHNIQUES, AS WELL AS EXPLORING NEWER TOPICS SUCH AS QUESTION ANSWERING AND THE USE OF DATA MINING AND MACHINE LEARNING. OF THE 42 SUBMISSIONS RECEIVED FOR THIS EDITION, 12 HAVE BEEN SELECTED FOR PUBLICATION AS FULL PAPERS AND 13 AS SHORT PAPERS, WITH AN ACCEPTANCE RATE OF AROUND 59%. THE PAPERS ADDRESS A WIDE RANGE OF TOPICS IN ARTIFICIAL INTELLIGENCE AND LAW, SUCH AS ARGUMENTATION, NORMS, EVIDENCE, BELIEF REVISION, CITATIONS, CASE-BASED REASONING AND ONTOLOGIES. DIVERSE TECHNIQUES SUCH AS INFORMATION RETRIEVAL AND EXTRACTION, MACHINE LEARNING, SEMANTIC WEB, AND NETWORK ANALYSIS WERE APPLIED, AMONG OTHERS, AND TEXTUAL SOURCES INCLUDE LEGAL CASES, BAR EXAMINATIONS, AND LEGISLATIVE/REGULATORY DOCUMENTS. THE BOOK WILL BE OF INTEREST TO ALL THOSE WORKING IN THE LEGAL SYSTEM WHO WISH TO KEEP ABREAST OF THE LATEST DEVELOPMENTS IN INFORMATION SYSTEMS.

**DATA LAKE FOR ENTERPRISES** - TOMCY JOHN 2017-05-31

A PRACTICAL GUIDE TO IMPLEMENTING YOUR ENTERPRISE DATA LAKE USING LAMBDA ARCHITECTURE AS THE BASE ABOUT THIS BOOK BUILD A FULL-FLEDGED DATA LAKE FOR YOUR ORGANIZATION WITH POPULAR BIG DATA TECHNOLOGIES USING THE LAMBDA ARCHITECTURE AS THE BASE DELVE INTO THE BIG DATA TECHNOLOGIES REQUIRED TO MEET MODERN DAY BUSINESS STRATEGIES A HIGHLY PRACTICAL GUIDE TO IMPLEMENTING ENTERPRISE DATA LAKES WITH LOTS OF EXAMPLES AND REAL-WORLD USE-CASES WHO THIS BOOK IS FOR JAVA DEVELOPERS AND ARCHITECTS WHO WOULD LIKE TO IMPLEMENT A DATA LAKE FOR THEIR ENTERPRISE WILL FIND THIS BOOK USEFUL. IF YOU WANT TO GET HANDS-ON EXPERIENCE WITH THE LAMBDA ARCHITECTURE AND BIG DATA TECHNOLOGIES BY IMPLEMENTING A PRACTICAL SOLUTION USING THESE TECHNOLOGIES, THIS BOOK WILL ALSO HELP YOU. WHAT YOU WILL LEARN BUILD AN ENTERPRISE-LEVEL DATA LAKE USING THE RELEVANT BIG DATA TECHNOLOGIES UNDERSTAND THE CORE OF THE LAMBDA ARCHITECTURE AND HOW TO APPLY IT IN AN ENTERPRISE LEARN THE TECHNICAL DETAILS AROUND SQOOP AND ITS FUNCTIONALITIES INTEGRATE KAFKA WITH HADOOP COMPONENTS TO ACQUIRE ENTERPRISE DATA USE FLUME WITH STREAMING TECHNOLOGIES FOR STREAM-BASED PROCESSING UNDERSTAND STREAM-BASED PROCESSING WITH REFERENCE TO APACHE SPARK STREAMING INCORPORATE HADOOP COMPONENTS AND KNOW THE ADVANTAGES THEY PROVIDE FOR ENTERPRISE DATA LAKES BUILD FAST, STREAMING, AND HIGH-PERFORMANCE APPLICATIONS USING ELASTICSEARCH MAKE YOUR DATA INGESTION PROCESS CONSISTENT ACROSS VARIOUS DATA FORMATS WITH CONFIGURABILITY PROCESS YOUR DATA TO DERIVE INTELLIGENCE USING MACHINE LEARNING ALGORITHMS IN DETAIL THE TERM "DATA LAKE" HAS RECENTLY EMERGED AS A PROMINENT TERM IN THE BIG DATA INDUSTRY. DATA SCIENTISTS CAN MAKE USE OF IT IN DERIVING MEANINGFUL INSIGHTS THAT CAN BE USED BY BUSINESSES TO REDEFINE OR TRANSFORM THE WAY THEY OPERATE. LAMBDA ARCHITECTURE IS ALSO EMERGING AS ONE OF THE VERY EMINENT PATTERNS IN THE BIG DATA LANDSCAPE, AS IT NOT ONLY HELPS TO DERIVE USEFUL INFORMATION FROM HISTORICAL DATA BUT ALSO CORRELATES REAL-TIME DATA TO ENABLE BUSINESS TO TAKE CRITICAL DECISIONS. THIS BOOK TRIES TO BRING THESE TWO IMPORTANT ASPECTS — DATA LAKE AND LAMBDA ARCHITECTURE—TOGETHER. THIS BOOK IS DIVIDED INTO THREE MAIN SECTIONS. THE FIRST INTRODUCES YOU TO THE CONCEPT OF DATA LAKES, THE IMPORTANCE OF DATA LAKES IN ENTERPRISES, AND GETTING YOU UP-TO-SPEED WITH THE LAMBDA ARCHITECTURE. THE SECOND SECTION DELVES INTO THE PRINCIPAL COMPONENTS OF

BUILDING A DATA LAKE USING THE LAMBDA ARCHITECTURE. IT INTRODUCES YOU TO POPULAR BIG DATA TECHNOLOGIES SUCH AS APACHE HADOOP, SPARK, SQOOP, FLUME, AND ELASTICSEARCH. THE THIRD SECTION IS A HIGHLY PRACTICAL DEMONSTRATION OF PUTTING IT ALL TOGETHER, AND SHOWS YOU HOW AN ENTERPRISE DATA LAKE CAN BE IMPLEMENTED, ALONG WITH SEVERAL REAL-WORLD USE-CASES. IT ALSO SHOWS YOU HOW OTHER PERIPHERAL COMPONENTS CAN BE ADDED TO THE LAKE TO MAKE IT MORE EFFICIENT. BY THE END OF THIS BOOK, YOU WILL BE ABLE TO CHOOSE THE RIGHT BIG DATA TECHNOLOGIES USING THE LAMBDA ARCHITECTURAL PATTERNS TO BUILD YOUR ENTERPRISE DATA LAKE. STYLE AND APPROACH THE BOOK TAKES A PRAGMATIC APPROACH, SHOWING WAYS TO LEVERAGE BIG DATA TECHNOLOGIES AND LAMBDA ARCHITECTURE TO BUILD AN ENTERPRISE-LEVEL DATA LAKE.

**IBM SPECTRUM DISCOVER: METADATA MANAGEMENT FOR DEEP INSIGHT OF UNSTRUCTURED STORAGE** - JOSEPH DAIN 2019-10-01

THIS IBM® REDPAPER PUBLICATION PROVIDES A COMPREHENSIVE OVERVIEW OF THE IBM SPECTRUM® DISCOVER METADATA MANAGEMENT SOFTWARE PLATFORM. WE GIVE A DETAILED EXPLANATION OF HOW THE PRODUCT CREATES, COLLECTS, AND ANALYZES METADATA. SEVERAL IN-DEPTH USE CASES ARE USED THAT SHOW EXAMPLES OF ANALYTICS, GOVERNANCE, AND OPTIMIZATION. WE ALSO PROVIDE STEP-BY-STEP INFORMATION TO INSTALL AND SET UP THE IBM SPECTRUM DISCOVER TRIAL ENVIRONMENT. MORE THAN 80% OF ALL DATA THAT IS COLLECTED BY ORGANIZATIONS IS NOT IN A STANDARD RELATIONAL DATABASE. INSTEAD, IT IS TRAPPED IN UNSTRUCTURED DOCUMENTS, SOCIAL MEDIA POSTS, MACHINE LOGS, AND SO ON. MANY ORGANIZATIONS FACE SIGNIFICANT CHALLENGES TO MANAGE THIS DELUGE OF UNSTRUCTURED DATA SUCH AS: PINPOINTING AND ACTIVATING RELEVANT DATA FOR LARGE-SCALE ANALYTICS LACKING THE FINE-GRAINED VISIBILITY THAT IS NEEDED TO MAP DATA TO BUSINESS PRIORITIES REMOVING REDUNDANT, OBSOLETE, AND TRIVIAL (ROT) DATA IDENTIFYING AND CLASSIFYING SENSITIVE DATA IBM SPECTRUM DISCOVER IS A MODERN METADATA MANAGEMENT SOFTWARE THAT PROVIDES DATA INSIGHT FOR PETABYTE-SCALE FILE AND OBJECT STORAGE, STORAGE ON PREMISES, AND IN THE CLOUD. THIS SOFTWARE ENABLES ORGANIZATIONS TO MAKE BETTER BUSINESS DECISIONS AND GAIN AND MAINTAIN A COMPETITIVE ADVANTAGE. IBM SPECTRUM DISCOVER PROVIDES A RICH METADATA LAYER THAT ENABLES STORAGE ADMINISTRATORS, DATA STEWARDS, AND DATA SCIENTISTS TO EFFICIENTLY MANAGE, CLASSIFY, AND GAIN INSIGHTS FROM MASSIVE AMOUNTS OF UNSTRUCTURED DATA. IT IMPROVES STORAGE ECONOMICS, HELPS MITIGATE RISK, AND ACCELERATES LARGE-SCALE ANALYTICS TO CREATE COMPETITIVE ADVANTAGE AND SPEED CRITICAL RESEARCH.

**INTRODUCING DATA SCIENCE** - DAVY CIELEN 2016-05-02

SUMMARY INTRODUCING DATA SCIENCE TEACHES YOU HOW TO ACCOMPLISH THE FUNDAMENTAL TASKS THAT OCCUPY DATA SCIENTISTS. USING THE PYTHON LANGUAGE AND COMMON PYTHON LIBRARIES, YOU'LL EXPERIENCE FIRSHTHAND THE CHALLENGES OF DEALING WITH DATA AT SCALE AND GAIN A SOLID FOUNDATION IN DATA SCIENCE. PURCHASE OF THE PRINT BOOK INCLUDES A FREE eBook IN PDF, KINDLE, AND ePub FORMATS FROM MANNING PUBLICATIONS. ABOUT THE TECHNOLOGY MANY COMPANIES NEED DEVELOPERS WITH DATA SCIENCE SKILLS TO WORK ON PROJECTS RANGING FROM SOCIAL MEDIA MARKETING TO MACHINE LEARNING. DISCOVERING WHAT YOU NEED TO LEARN TO BEGIN A CAREER AS A DATA SCIENTIST CAN SEEM BEWILDERING. THIS BOOK IS DESIGNED TO HELP YOU GET STARTED. ABOUT THE BOOK INTRODUCING DATA SCIENCEINTRODUCING DATA SCIENCE EXPLAINS VITAL DATA SCIENCE CONCEPTS AND TEACHES YOU HOW TO ACCOMPLISH THE FUNDAMENTAL TASKS THAT OCCUPY DATA SCIENTISTS. YOU'LL EXPLORE DATA VISUALIZATION, GRAPH DATABASES, THE USE OF NoSQL, AND THE DATA SCIENCE PROCESS. YOU'LL USE THE PYTHON LANGUAGE AND COMMON PYTHON LIBRARIES AS YOU EXPERIENCE FIRSHTHAND THE CHALLENGES OF DEALING WITH DATA AT SCALE. DISCOVER HOW PYTHON ALLOWS YOU TO GAIN INSIGHTS FROM DATA SETS SO BIG THAT THEY NEED TO BE STORED ON MULTIPLE MACHINES, OR FROM DATA MOVING SO QUICKLY THAT NO SINGLE MACHINE CAN HANDLE IT. THIS BOOK GIVES YOU HANDS-ON EXPERIENCE WITH THE MOST POPULAR PYTHON DATA SCIENCE LIBRARIES, SCIKIT-LEARN AND STATSModels. AFTER READING THIS BOOK, YOU'LL HAVE THE SOLID FOUNDATION YOU NEED TO START A CAREER IN DATA SCIENCE. WHAT'S INSIDE HANDLING LARGE DATA INTRODUCTION TO MACHINE LEARNING USING PYTHON TO WORK WITH DATA WRITING DATA SCIENCE ALGORITHMS ABOUT THE READER THIS BOOK ASSUMES YOU'RE COMFORTABLE READING CODE IN PYTHON OR A SIMILAR LANGUAGE, SUCH AS C, RUBY, OR JAVASCRIPT. NO PRIOR EXPERIENCE WITH DATA SCIENCE IS REQUIRED. ABOUT THE AUTHORS DAVY CIELEN, ARNO D. B. MEYSMAN, AND MOHAMED ALI ARE THE FOUNDERS AND MANAGING PARTNERS OF OPTIMATELY AND MAITON, WHERE THEY FOCUS ON DEVELOPING DATA SCIENCE PROJECTS AND SOLUTIONS IN VARIOUS SECTORS. TABLE OF CONTENTS DATA SCIENCE IN A BIG DATA WORLD THE DATA SCIENCE PROCESS MACHINE LEARNING HANDLING LARGE DATA ON A SINGLE COMPUTER FIRST STEPS IN BIG DATA JOIN THE NoSQL MOVEMENT THE RISE OF GRAPH DATABASES TEXT MINING AND TEXT ANALYTICS DATA VISUALIZATION TO THE END USER

**XQUERY KICK START** - JAMES MCGOVERN 2004

"XQUERY KICK START" DELIVERS A CONCISE INTRODUCTION TO THE XQUERY STANDARD, AND USEFUL IMPLEMENTATION ADVICE FOR DEVELOPERS NEEDING TO PUT IT INTO PRACTICE. THE BOOK STARTS BY EXPLAINING THE ROLE OF XQUERY IN THE XML FAMILY OF SPECIFICATIONS, AND ITS RELATIONSHIP WITH XPATH. THE AUTHORS THEN EXPLAIN THE SPECIFICATION IN DETAIL, DESCRIBING THE SEMANTICS AND DATA MODEL, BEFORE MOVING TO EXAMPLES USING XQUERY TO MANIPULATE XML DATABASES AND DOCUMENT STORAGE SYSTEMS. LATER CHAPTERS DISCUSS JAVA IMPLEMENTATIONS OF XQUERY AND DEVELOPMENT TOOLS THAT FACILITATE THE DEVELOPMENT OF WEB SITES WITH XQUERY. THIS BOOK IS UP TO DATE WITH THE LATEST XQUERY SPECIFICATIONS, AND INCLUDES COVERAGE OF NEW FEATURES FOR EXTENDING THE XQUERY LANGUAGE.

**THE FUTURE OF EMAIL ARCHIVES** - TASK FORCE ON TECHNICAL APPROACHES TO EMAIL ARCHIVES 2018-09

*TEXT PROCESSING WITH GATE* - HAMISH CUNNINGHAM 2011

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/](http://gate.ac.uk/).

*DATABASE AND EXPERT SYSTEMS APPLICATIONS* - MOURAD ELLOUMI 2018-08-06

THIS VOLUME CONSTITUTES THE REFEREED PROCEEDINGS OF THE THREE WORKSHOPS HELD AT THE 29TH INTERNATIONAL CONFERENCE ON DATABASE AND EXPERT SYSTEMS APPLICATIONS, DEXA 2018, HELD IN REGENSBURG, GERMANY, IN SEPTEMBER 2018: THE THIRD INTERNATIONAL WORKSHOP ON BIG DATA MANAGEMENT IN CLOUD SYSTEMS, BDMICS 2018, THE 9TH INTERNATIONAL WORKSHOP ON BIOLOGICAL KNOWLEDGE DISCOVERY FROM DATA, BLOKDD, AND THE 15TH INTERNATIONAL WORKSHOP ON TECHNOLOGIES FOR INFORMATION RETRIEVAL, TIR. THE 25 REVISED FULL PAPERS WERE CAREFULLY REVIEWED AND SELECTED FROM 33 SUBMISSIONS. THE PAPERS DISCUSS A RANGE OF TOPICS INCLUDING: PARALLEL DATA MANAGEMENT SYSTEMS, CONSISTENCY AND PRIVACY CLOUD COMPUTING AND GRAPH QUERIES, WEB AND DOMAIN CORPORA, NLP APPLICATIONS, SOCIAL MEDIA AND PERSONALIZATION

*PRACTICAL WEB SCRAPING FOR DATA SCIENCE* - SEPPE VANDEN BROUCKE 2018-04-18

THIS BOOK PROVIDES A COMPLETE AND MODERN GUIDE TO WEB SCRAPING, USING PYTHON AS THE PROGRAMMING LANGUAGE, WITHOUT GLOSSING OVER IMPORTANT DETAILS OR BEST PRACTICES. WRITTEN WITH A DATA SCIENCE AUDIENCE IN MIND, THE BOOK EXPLORES BOTH SCRAPING AND THE LARGER CONTEXT OF WEB TECHNOLOGIES IN WHICH IT OPERATES, TO ENSURE FULL UNDERSTANDING. THE AUTHORS RECOMMEND WEB SCRAPING AS A POWERFUL TOOL FOR ANY DATA SCIENTIST'S ARSENAL, AS MANY DATA SCIENCE PROJECTS START BY OBTAINING AN APPROPRIATE DATA SET. STARTING WITH A BRIEF OVERVIEW ON SCRAPING AND REAL-LIFE USE CASES, THE AUTHORS EXPLORE THE CORE CONCEPTS OF HTTP, HTML, AND CSS TO PROVIDE A SOLID FOUNDATION. ALONG WITH A QUICK PYTHON PRIMER, THEY COVER SELENIUM FOR JAVASCRIPT-HEAVY SITES, AND WEB CRAWLING IN DETAIL. THE BOOK FINISHES WITH A RECAP OF BEST PRACTICES AND A COLLECTION OF EXAMPLES THAT BRING TOGETHER EVERYTHING YOU'VE LEARNED AND ILLUSTRATE VARIOUS DATA SCIENCE USE CASES. WHAT YOU'LL LEARN LEVERAGE WELL-ESTABLISHED BEST PRACTICES AND COMMONLY-USED PYTHON PACKAGES HANDLE TODAY'S WEB, INCLUDING JAVASCRIPT, COOKIES, AND COMMON WEB SCRAPING MITIGATION TECHNIQUES UNDERSTAND THE MANAGERIAL AND LEGAL CONCERNS REGARDING WEB SCRAPING WHO THIS BOOK IS FOR A DATA SCIENCE ORIENTED AUDIENCE THAT IS PROBABLY ALREADY FAMILIAR WITH PYTHON OR ANOTHER PROGRAMMING LANGUAGE OR ANALYTICAL TOOLKIT (R, SAS, SPSS, ETC). STUDENTS OR INSTRUCTORS IN UNIVERSITY COURSES MAY ALSO BENEFIT. READERS UNFAMILIAR WITH PYTHON WILL APPRECIATE A QUICK PYTHON PRIMER IN CHAPTER 1 TO CATCH UP WITH THE BASICS AND PROVIDE POINTERS TO OTHER GUIDES AS WELL.

*PRODUCT LIFECYCLE MANAGEMENT IN THE DIGITAL TWIN ERA* - CLEMENT FORTIN 2020-02-28

THIS BOOK CONSTITUTES THE REFEREED POST-CONFERENCE PROCEEDINGS OF THE 16TH IFIP WG 5.1 INTERNATIONAL CONFERENCE ON PRODUCT LIFECYCLE MANAGEMENT, PLM 2019, HELD IN MOSCOW, RUSSIA, IN JULY 2019. THE 38 REVISED FULL PAPERS PRESENTED WERE CAREFULLY REVIEWED AND SELECTED FROM 63 SUBMISSIONS. THE PAPERS ARE ORGANIZED IN THE FOLLOWING TOPICAL SECTIONS: 3D MODELLING AND DATA STRUCTURES; PLM MATURITY AND INDUSTRY 4.0; ONTOLOGIES AND SEMANTICS; PLM AND CONCEPTUAL DESIGN; KNOWLEDGE AND CHANGE MANAGEMENT; IoT AND PLM; INTEGRATING MANUFACTURING REALITIES; AND INTEGRATION OF IN-SERVICE AND OPERATION.

*ICT ANALYSIS AND APPLICATIONS* - SIMON FONG 2020-12-15

THIS BOOK PROPOSES NEW TECHNOLOGIES AND DISCUSSES FUTURE SOLUTIONS FOR ICT DESIGN INFRASTRUCTURES, AS REFLECTED IN HIGH-QUALITY PAPERS PRESENTED AT THE 5TH INTERNATIONAL CONFERENCE ON ICT FOR SUSTAINABLE DEVELOPMENT (ICT4SD 2020), HELD IN GOA, INDIA, ON 23-24 JULY 2020. THE CONFERENCE PROVIDED A VALUABLE FORUM FOR CUTTING-EDGE RESEARCH DISCUSSIONS AMONG PIONEERING RESEARCHERS, SCIENTISTS, INDUSTRIAL ENGINEERS, AND STUDENTS FROM ALL AROUND THE WORLD. BRINGING TOGETHER EXPERTS FROM DIFFERENT COUNTRIES, THE BOOK EXPLORES A RANGE OF CENTRAL ISSUES FROM AN INTERNATIONAL PERSPECTIVE.

*BUILDING DIGITAL EXPERIENCE PLATFORMS* - SHAILESH KUMAR SHIVAKUMAR 2019-01-31

USE DIGITAL EXPERIENCE PLATFORMS (DXP) TO IMPROVE YOUR DEVELOPMENT PRODUCTIVITY AND RELEASE TIMELINES. LEVERAGE THE PRE-INTEGRATED FEATURE SETS OF DXPs IN YOUR ORGANIZATION'S DIGITAL TRANSFORMATION JOURNEY TO QUICKLY DEVELOP A PERSONALIZED, SECURE, AND ROBUST ENTERPRISE PLATFORM. IN THIS BOOK THE AUTHORS EXAMINE VARIOUS FEATURES OF DXPs AND PROVIDE RICH INSIGHTS INTO BUILDING EACH LAYER IN A DIGITAL PLATFORM. PROVEN BEST PRACTICES ARE PRESENTED WITH EXAMPLES FOR DESIGNING AND BUILDING LAYERS. A SPECIAL FOCUS IS PROVIDED ON SECURITY AND QUALITY ATTRIBUTES NEEDED FOR BUSINESS-CRITICAL ENTERPRISE APPLICATIONS. THE AUTHORS COVER MODERN AND EMERGING DIGITAL TRENDS SUCH AS BLOCKCHAIN, IoT, CONTAINERS, CHATBOTS, ARTIFICIAL INTELLIGENCE, AND MORE. THE BOOK IS DIVIDED INTO FIVE PARTS RELATED TO REQUIREMENTS/DESIGN, DEVELOPMENT, SECURITY, INFRASTRUCTURE, AND CASE STUDY. THE AUTHORS EMPLOY PROVEN REAL-WORLD METHODS, BEST PRACTICES, AND SECURITY AND INTEGRATION TECHNIQUES DERIVED FROM THEIR RICH EXPERIENCE. AN ELABORATE DIGITAL TRANSFORMATION CASE STUDY FOR A BANKING APPLICATION IS INCLUDED. WHAT YOU'LL LEARN DEVELOP A DIGITAL EXPERIENCE PLATFORM FROM END TO END UNDERSTAND BEST PRACTICES AND PROVEN METHODS FOR DESIGNING OVERALL ARCHITECTURE, USER INTERFACE AND INTEGRATION COMPONENTS, SECURITY, AND INFRASTRUCTURE STUDY REAL-WORLD CASES, INCLUDING AN ELABORATE DIGITAL TRANSFORMATION BUILDING AN ENTERPRISE PLATFORM FOR A BANKING APPLICATION KNOW THE OPEN SOURCE TOOLS AND

TECHNOLOGY FRAMEWORKS THAT CAN BE USED TO BUILD DXPs WHO THIS BOOK IS FOR WEB DEVELOPERS, FULL STACK DEVELOPERS, DIGITAL ENTHUSIASTS, DIGITAL PROJECT MANAGERS, AND ARCHITECTS

*SOLR COOKBOOK - THIRD EDITION* - RAFAL KUC 2015-01-23

THIS BOOK IS FOR INTERMEDIATE SOLR DEVELOPERS WHO ARE WILLING TO LEARN AND IMPLEMENT PRO-LEVEL PRACTICES, TECHNIQUES, AND SOLUTIONS. THIS EDITION WILL SPECIFICALLY APPEAL TO DEVELOPERS WHO WISH TO QUICKLY GET TO GRIPS WITH THE CHANGES AND NEW FEATURES OF APACHE SOLR 5.

*MACHINE LEARNING WITH TENSORFLOW, SECOND EDITION* - MATTMANN A. CHRIS 2021-02-02

UPDATED WITH NEW CODE, NEW PROJECTS, AND NEW CHAPTERS, MACHINE LEARNING WITH TENSORFLOW, SECOND EDITION GIVES READERS A SOLID FOUNDATION IN MACHINE-LEARNING CONCEPTS AND THE TENSORFLOW LIBRARY. SUMMARY UPDATED WITH NEW CODE, NEW PROJECTS, AND NEW CHAPTERS, MACHINE LEARNING WITH TENSORFLOW, SECOND EDITION GIVES READERS A SOLID FOUNDATION IN MACHINE-LEARNING CONCEPTS AND THE TENSORFLOW LIBRARY. WRITTEN BY NASA JPL DEPUTY CTO AND PRINCIPAL DATA SCIENTIST CHRIS MATTMANN, ALL EXAMPLES ARE ACCOMPANIED BY DOWNLOADABLE JUPYTER NOTEBOOKS FOR A HANDS-ON EXPERIENCE CODING TENSORFLOW WITH PYTHON. NEW AND REVISED CONTENT EXPANDS COVERAGE OF CORE MACHINE LEARNING ALGORITHMS, AND ADVANCEMENTS IN NEURAL NETWORKS SUCH AS VGG-FACE FACIAL IDENTIFICATION CLASSIFIERS AND DEEP SPEECH CLASSIFIERS. PURCHASE OF THE PRINT BOOK INCLUDES A FREE eBook IN PDF, KINDLE, AND ePub FORMATS FROM MANNING PUBLICATIONS. ABOUT THE TECHNOLOGY SUPERCHARGE YOUR DATA ANALYSIS WITH MACHINE LEARNING! ML ALGORITHMS AUTOMATICALLY IMPROVE AS THEY PROCESS DATA, SO RESULTS GET BETTER OVER TIME. YOU DON'T HAVE TO BE A MATHEMATICIAN TO USE ML: TOOLS LIKE GOOGLE'S TENSORFLOW LIBRARY HELP WITH COMPLEX CALCULATIONS SO YOU CAN FOCUS ON GETTING THE ANSWERS YOU NEED. ABOUT THE BOOK MACHINE LEARNING WITH TENSORFLOW, SECOND EDITION IS A FULLY REVISED GUIDE TO BUILDING MACHINE LEARNING MODELS USING PYTHON AND TENSORFLOW. YOU'LL APPLY CORE ML CONCEPTS TO REAL-WORLD CHALLENGES, SUCH AS SENTIMENT ANALYSIS, TEXT CLASSIFICATION, AND IMAGE RECOGNITION. HANDS-ON EXAMPLES ILLUSTRATE NEURAL NETWORK TECHNIQUES FOR DEEP SPEECH PROCESSING, FACIAL IDENTIFICATION, AND AUTO-ENCODING WITH CIFAR-10. WHAT'S INSIDE MACHINE LEARNING WITH TENSORFLOW CHOOSING THE BEST ML APPROACHES VISUALIZING ALGORITHMS WITH TENSORBOARD SHARING RESULTS WITH COLLABORATORS RUNNING MODELS IN DOCKER ABOUT THE READER REQUIRES INTERMEDIATE PYTHON SKILLS AND KNOWLEDGE OF GENERAL ALGEBRAIC CONCEPTS LIKE VECTORS AND MATRICES. EXAMPLES USE THE SUPER-STABLE 1.15.X BRANCH OF TENSORFLOW AND TENSORFLOW 2.X. ABOUT THE AUTHOR CHRIS MATTMANN IS THE DIVISION MANAGER OF THE ARTIFICIAL INTELLIGENCE, ANALYTICS, AND INNOVATION ORGANIZATION AT NASA JET PROPULSION LAB. THE FIRST EDITION OF THIS BOOK WAS WRITTEN BY NISHANT SHUKLA WITH KENNETH FRICKLAS. TABLE OF CONTENTS PART 1 - YOUR MACHINE-LEARNING RIG 1 A MACHINE-LEARNING ODYSSEY 2 TENSORFLOW ESSENTIALS PART 2 - CORE LEARNING ALGORITHMS 3 LINEAR REGRESSION AND BEYOND 4 USING REGRESSION FOR CALL-CENTER VOLUME PREDICTION 5 A GENTLE INTRODUCTION TO CLASSIFICATION 6 SENTIMENT CLASSIFICATION: LARGE MOVIE-REVIEW DATASET 7 AUTOMATICALLY CLUSTERING DATA 8 INFERRING USER ACTIVITY FROM ANDROID ACCELEROMETER DATA 9 HIDDEN MARKOV MODELS 10 PART-OF-SPEECH TAGGING AND WORD-SENSE DISAMBIGUATION PART 3 - THE NEURAL NETWORK PARADIGM 11 A PEEK INTO AUTOENCODERS 12 APPLYING AUTOENCODERS: THE CIFAR-10 IMAGE DATASET 13 REINFORCEMENT LEARNING 14 CONVOLUTIONAL NEURAL NETWORKS 15 BUILDING A REAL-WORLD CNN: VGG-FACE AD VGG-FACE LITE 16 RECURRENT NEURAL NETWORKS 17 LSTMS AND AUTOMATIC SPEECH RECOGNITION 18 SEQUENCE-TO-SEQUENCE MODELS FOR CHATBOTS 19 UTILITY LANDSCAPE HEADLINESE - INGRID M[R] RDH 1980

*SOLR IN ACTION* - TIMOTHY POTTER 2014-03-25

SUMMARY SOLR IN ACTION IS A COMPREHENSIVE GUIDE TO IMPLEMENTING SCALABLE SEARCH USING APACHE SOLR. THIS CLEARLY WRITTEN BOOK WALKS YOU THROUGH WELL-DOCUMENTED EXAMPLES RANGING FROM BASIC KEYWORD SEARCHING TO SCALING A SYSTEM FOR BILLIONS OF DOCUMENTS AND QUERIES. IT WILL GIVE YOU A DEEP UNDERSTANDING OF HOW TO IMPLEMENT CORE SOLR CAPABILITIES. ABOUT THE BOOK WHETHER YOU'RE HANDLING BIG (OR SMALL) DATA, MANAGING DOCUMENTS, OR BUILDING A WEBSITE, IT IS IMPORTANT TO BE ABLE TO QUICKLY SEARCH THROUGH YOUR CONTENT AND DISCOVER MEANING IN IT. APACHE SOLR IS YOUR TOOL: A READY-TO-DEPLOY, LUCENE-BASED, OPEN SOURCE, FULL-TEXT SEARCH ENGINE. SOLR CAN SCALE ACROSS MANY SERVERS TO ENABLE REAL-TIME QUERIES AND DATA ANALYTICS ACROSS BILLIONS OF DOCUMENTS. SOLR IN ACTION TEACHES YOU TO IMPLEMENT SCALABLE SEARCH USING APACHE SOLR. THIS EASY-TO-READ GUIDE BALANCES CONCEPTUAL DISCUSSIONS WITH PRACTICAL EXAMPLES TO SHOW YOU HOW TO IMPLEMENT ALL OF SOLR'S CORE CAPABILITIES. YOU'LL MASTER TOPICS LIKE TEXT ANALYSIS, FACETED SEARCH, HIT HIGHLIGHTING, RESULT GROUPING, QUERY SUGGESTIONS, MULTILINGUAL SEARCH, ADVANCED GEOSPATIAL AND DATA OPERATIONS, AND RELEVANCY TUNING. THIS BOOK ASSUMES BASIC KNOWLEDGE OF JAVA AND STANDARD DATABASE TECHNOLOGY. NO PRIOR KNOWLEDGE OF SOLR OR LUCENE IS REQUIRED. PURCHASE OF THE PRINT BOOK INCLUDES A FREE eBook IN PDF, KINDLE, AND ePub FORMATS FROM MANNING PUBLICATIONS. WHAT'S INSIDE HOW TO SCALE SOLR FOR BIG DATA RICH REAL-WORLD EXAMPLES SOLR AS A NoSQL DATA STORE ADVANCED MULTILINGUAL, DATA, AND RELEVANCY TRICKS COVERAGE OF VERSIONS THROUGH SOLR 4.7 ABOUT THE AUTHORS TREY GRAINGER IS A DIRECTOR OF ENGINEERING AT CAREERBUILDER. TIMOTHY POTTER IS A SENIOR MEMBER OF THE ENGINEERING TEAM AT LUCIDWORKS. THE AUTHORS WORK ON THE SCALABILITY AND RELIABILITY OF SOLR, AS WELL AS ON RECOMMENDATION ENGINE AND BIG DATA ANALYTICS TECHNOLOGIES. TABLE OF CONTENTS PART 1 MEET SOLR INTRODUCTION TO SOLR GETTING TO KNOW SOLR KEY SOLR CONCEPTS CONFIGURING SOLR INDEXING TEXT ANALYSIS PART 2 CORE SOLR CAPABILITIES PERFORMING QUERIES AND HANDLING RESULTS FACETED SEARCH HIT HIGHLIGHTING QUERY SUGGESTIONS RESULT GROUPING/FIELD COLLAPSING TAKING SOLR TO PRODUCTION PART 3 TAKING SOLR TO THE NEXT LEVEL SOLRCloud MULTILINGUAL SEARCH COMPLEX QUERY OPERATIONS MASTERING RELEVANCY



## GUIDE TO CLOUD COMPUTING - RICHARD HILL 2012-11-28

THIS BOOK DESCRIBES THE LANDSCAPE OF CLOUD COMPUTING FROM FIRST PRINCIPLES, LEADING THE READER STEP-BY-STEP THROUGH THE PROCESS OF BUILDING AND CONFIGURING A CLOUD ENVIRONMENT. THE BOOK NOT ONLY CONSIDERS THE TECHNOLOGIES FOR DESIGNING AND CREATING CLOUD COMPUTING PLATFORMS, BUT ALSO THE BUSINESS MODELS AND FRAMEWORKS IN REAL-WORLD IMPLEMENTATION OF CLOUD PLATFORMS. EMPHASIS IS PLACED ON “LEARNING BY DOING,” AND READERS ARE ENCOURAGED TO EXPERIMENT WITH A RANGE OF DIFFERENT TOOLS AND APPROACHES. TOPICS AND FEATURES: INCLUDES REVIEW QUESTIONS, HANDS-ON EXERCISES, STUDY ACTIVITIES AND DISCUSSION TOPICS THROUGHOUT THE TEXT; DEMONSTRATES THE APPROACHES USED TO BUILD CLOUD COMPUTING INFRASTRUCTURES; REVIEWS THE SOCIAL, ECONOMIC, AND POLITICAL ASPECTS OF THE ON-GOING GROWTH IN CLOUD COMPUTING USE; DISCUSSES LEGAL AND SECURITY CONCERNS IN CLOUD COMPUTING; EXAMINES TECHNIQUES FOR THE APPRAISAL OF FINANCIAL INVESTMENT INTO CLOUD COMPUTING; IDENTIFIES AREAS FOR FURTHER RESEARCH WITHIN THIS RAPIDLY-MOVING FIELD.

## TAMING TEXT - GRANT INGERSOLL 2012-12-20

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 COMMITTEE MEMBER, 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, SCHOOL OF INFORMATION SCIENCE, 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

## HADOOP: THE DEFINITIVE GUIDE - TOM WHITE 2012-05-10

READY TO UNLOCK THE POWER OF YOUR DATA? WITH THIS COMPREHENSIVE GUIDE, YOU’LL LEARN HOW TO BUILD AND MAINTAIN RELIABLE, SCALABLE, DISTRIBUTED SYSTEMS WITH APACHE HADOOP. THIS BOOK IS IDEAL FOR PROGRAMMERS LOOKING TO ANALYZE DATASETS OF ANY SIZE, AND FOR ADMINISTRATORS WHO WANT TO SET UP AND RUN HADOOP CLUSTERS. YOU’LL FIND ILLUMINATING CASE STUDIES THAT DEMONSTRATE HOW HADOOP IS USED TO SOLVE SPECIFIC PROBLEMS. THIS THIRD EDITION COVERS RECENT CHANGES TO HADOOP, INCLUDING MATERIAL ON THE NEW MAPREDUCE API, AS WELL AS MAPREDUCE 2 AND ITS MORE FLEXIBLE EXECUTION MODEL (YARN). STORE LARGE DATASETS WITH THE HADOOP DISTRIBUTED FILE SYSTEM (HDFS) RUN DISTRIBUTED COMPUTATIONS WITH MAPREDUCE USE HADOOP’S DATA AND I/O BUILDING BLOCKS FOR COMPRESSION, DATA INTEGRITY, SERIALIZATION (INCLUDING AVRO), AND PERSISTENCE DISCOVER COMMON PITFALLS AND ADVANCED FEATURES FOR WRITING REAL-WORLD MAPREDUCE PROGRAMS DESIGN, BUILD, AND ADMINISTER A DEDICATED HADOOP CLUSTER—OR RUN HADOOP IN THE CLOUD LOAD DATA FROM RELATIONAL DATABASES INTO HDFS, USING SQOOP PERFORM LARGE-SCALE DATA PROCESSING WITH THE PIG QUERY LANGUAGE ANALYZE DATASETS WITH HIVE, HADOOP’S DATA WAREHOUSING SYSTEM TAKE ADVANTAGE OF HBASE FOR STRUCTURED AND SEMI-STRUCTURED DATA, AND ZOOKEEPER FOR BUILDING DISTRIBUTED SYSTEMS

## ARTIFICIAL INTELLIGENCE IN DATA MINING - D. BINU 2021-02-17

ARTIFICIAL INTELLIGENCE IN DATA MINING: THEORIES AND APPLICATIONS OFFERS A COMPREHENSIVE INTRODUCTION TO DATA MINING THEORIES, RELEVANT AI TECHNIQUES, AND THEIR MANY REAL-WORLD APPLICATIONS. THIS BOOK IS WRITTEN BY EXPERIENCED ENGINEERS FOR ENGINEERS, BIOMEDICAL ENGINEERS, AND RESEARCHERS IN NEURAL NETWORKS, AS WELL AS COMPUTER SCIENTISTS WITH AN INTEREST IN THE AREA. PROVIDES COVERAGE OF THE FUNDAMENTALS OF ARTIFICIAL INTELLIGENCE AS APPLIED TO DATA MINING, INCLUDING COMPUTATIONAL INTELLIGENCE AND UNSUPERVISED LEARNING METHODS FOR DATA CLUSTERING PRESENTS COVERAGE OF KEY TOPICS SUCH AS HEURISTIC METHODS FOR DATA CLUSTERING, DEEP LEARNING METHODS FOR DATA CLASSIFICATION, AND NEURAL NETWORKS INCLUDES CASE STUDIES AND REAL-WORLD APPLICATIONS OF AI TECHNIQUES IN DATA MINING, FOR IMPROVED OUTCOMES IN CLINICAL DIAGNOSIS, SATELLITE DATA EXTRACTION, AGRICULTURE, SECURITY AND DEFENSE

## PROGRAMMING JOOMLA PLUGINS - JISSE REITSMA 2015-12-03

THIS BOOK - WRITTEN BY JISSE REITSMA - IS ALL ABOUT PLUGINS. AN ENTIRE BOOK ON THIS TOPIC? YES! JUST LIKE COMPONENTS AND MODULES, PLUGINS ALLOW YOU TO QUICKLY EXTEND JOOMLA. IT’S THE MOST FLEXIBLE WAY TO CHANGE THE BEHAVIOR OF JOOMLA AND

ADD COOL EXTRAS TO YOUR WEBSITE. IN THIS BOOK, YOU WILL LEARN HOW TO WRITE PLUGINS FROM SCRATCH. BOTH BEGINNER AND GURU WILL FIND TOPICS TO THEIR LIKING. ALL PLUGIN TYPES AND EVENTS ARE COVERED. THE BOOK ALSO HELPS WITH XML AND FRAMEWORK CALLS. IT DOES NOT STOP THERE: DESIGN PATTERNS, AUTOLOADERS, PHPUNIT, JFORM ADDITIONS ARE ALSO INCLUDED. EVERYTHING YOU WANT TO KNOW IS COVERED IN THE BOOK, MAKING IT THE REFERENCE ON DEVELOPING JOOMLA PLUGINS. ALL CODE CAN BE FOUND ON GITHUB, MAKING IT EASY FOR YOU TO START USING WHAT YOU LEARN.

## MASTERING APACHE SOLR 7.X - SANDEEP NAIR 2018-02-22

ACCELERATE YOUR ENTERPRISE SEARCH ENGINE AND BRING RELEVANCY IN YOUR SEARCH ANALYTICS KEY FEATURES A PRACTICAL GUIDE IN BUILDING EXPERTISE WITH INDEXING, FACETING, CLUSTERING AND PAGINATION MASTER THE MANAGEMENT AND ADMINISTRATION OF ENTERPRISE SEARCH APPLICATIONS AND SERVICES SEAMLESSLY HANDLE MULTIPLE DATA INPUTS SUCH AS JSON, XML, PDF, DOC, XLS, PPT, CSV AND MUCH MORE. BOOK DESCRIPTION APACHE SOLR IS THE ONLY STANDALONE ENTERPRISE SEARCH SERVER WITH A REST-LIKE APPLICATION INTERFACE. PROVIDING HIGHLY SCALABLE, DISTRIBUTED SEARCH AND INDEX REPLICATION FOR MANY OF THE WORLD’S LARGEST INTERNET SITES. TO BEGIN WITH, YOU WOULD BE INTRODUCED TO HOW YOU PERFORM FULL TEXT SEARCH, MULTIPLE FILTER SEARCH, PERFORM DYNAMIC CLUSTERING AND SO ON HELPING YOU TO BRUSH UP THE BASICS OF APACHE SOLR. YOU WILL ALSO EXPLORE THE NEW FEATURES AND ADVANCED OPTIONS RELEASED IN APACHE SOLR 7.X WHICH WILL GET YOU NUMEROUS PERFORMANCE ASPECTS AND MAKING DATA INVESTIGATION SIMPLER, EASIER AND POWERFUL. YOU WILL LEARN TO BUILD COMPLEX QUERIES, EXTENSIVE FILTERS AND HOW ARE THEY COMPILED IN YOUR SYSTEM TO BRING RELEVANCE IN YOUR SEARCH TOOLS. YOU WILL LEARN TO CARRY OUT SOLR SCORING, ELEMENTS AFFECTING THE DOCUMENT SCORE AND HOW YOU CAN OPTIMIZE OR TUNE THE SCORE FOR THE APPLICATION AT HAND. YOU WILL LEARN TO EXTRACT FEATURES OF DOCUMENTS, WRITING COMPLEX QUERIES IN RE-RANKING THE DOCUMENTS. YOU WILL ALSO LEARN ADVANCED OPTIONS HELPING YOU TO KNOW WHAT CONTENT IS INDEXED AND HOW THE EXTRACTED CONTENT IS INDEXED. THROUGHOUT THE BOOK, YOU WOULD GO THROUGH COMPLEX PROBLEMS WITH SOLUTIONS ALONG WITH VARIED APPROACHES TO TACKLE YOUR BUSINESS NEEDS. BY THE END OF THIS BOOK, YOU WILL GAIN ADVANCED PROFICIENCY TO BUILD OUT-OF-BOX SMART SEARCH SOLUTIONS FOR YOUR ENTERPRISE DEMANDS. WHAT YOU WILL LEARN DESIGN SCHEMA USING SCHEMA API TO ACCESS DATA IN THE DATABASE ADVANCE QUERYING AND FINE-TUNING TECHNIQUES FOR BETTER PERFORMANCE GET TO GRIPS WITH INDEXING USING CLIENT API SET UP A FAULT TOLERANT AND HIGHLY AVAILABLE SERVER WITH NEWER DISTRIBUTED CAPABILITIES, SOLR CLOUD EXPLORE APACHE TIKA TO UPLOAD DATA WITH SOLR CELL UNDERSTAND DIFFERENT DATA OPERATIONS THAT CAN BE DONE WHILE INDEXING MASTER ADVANCED QUERYING THROUGH VELOCITY SEARCH UI, FACETING AND QUERY RE-RANKING, PAGINATION AND SPATIAL SEARCH LEARN TO USE JAVASCRIPT, PYTHON, SOLR J AND RUBY FOR INTERACTING WITH SOLR WHO THIS BOOK IS FOR THE BOOK WOULD RIGHTLY APPEAL TO DEVELOPERS, SOFTWARE ENGINEERS, DATA ENGINEERS AND DATABASE ARCHITECTS WHO ARE BUILDING OR SEEKING TO BUILD ENTERPRISE-WIDE EFFECTIVE SEARCH ENGINES FOR BUSINESS INTELLIGENCE. PRIOR EXPERIENCE OF APACHE SOLR OR JAVA PROGRAMMING IS MUST TO TAKE THE BEST OF THIS BOOK.

## DIGITAL PRESERVATION METADATA FOR PRACTITIONERS - ANGELA DAPPERT 2016-12-19

THIS BOOK BEGINS WITH AN INTRODUCTION TO FUNDAMENTAL ISSUES RELATED TO DIGITAL PRESERVATION METADATA BEFORE PROCEEDING TO IN-DEPTH COVERAGE OF ISSUES CONCERNING ITS PRACTICAL USE AND IMPLEMENTATION. IT HELPS READERS TO UNDERSTAND WHICH OPTIONS NEED TO BE CONSIDERED IN SPECIFYING A DIGITAL PRESERVATION METADATA PROFILE TO ENSURE IT MATCHES THEIR INDIVIDUAL CONTENT TYPES, TECHNICAL INFRASTRUCTURE, AND ORGANIZATIONAL NEEDS. FURTHER, IT PROVIDES PRACTICAL GUIDANCE AND EXAMPLES, AND RAISES IMPORTANT QUESTIONS. IT DOES NOT PROVIDE FULL-FLEDGED IMPLEMENTATION SOLUTIONS, AS SUCH SOLUTIONS CAN, BY DEFINITION, ONLY BE SPECIFIC TO A GIVEN PRESERVATION CONTEXT. AS SUCH, THE BOOK EFFECTIVELY BRIDGES THE GAP BETWEEN THE FORMAL SPECIFICATIONS PROVIDED IN A STANDARD, SUCH AS THE PREMIS DATA DICTIONARY – A DE-FACTO STANDARD THAT DEFINES THE CORE METADATA REQUIRED BY MOST PRESERVATION REPOSITORIES – AND SPECIFIC IMPLEMENTATIONS. ANYBODY WHO NEEDS TO MANAGE DIGITAL ASSETS IN ANY FORM WITH THE INTENT OF PRESERVING THEM FOR AN INDEFINITE PERIOD OF TIME WILL FIND THIS BOOK A VALUABLE RESOURCE. THE PREMIS DATA DICTIONARY PROVIDES A DATA MODEL CONSISTING OF BASIC ENTITIES (OBJECTS, AGENTS, EVENTS AND RIGHTS) AND BASIC PROPERTIES (CALLED “SEMANTIC UNITS”) THAT DESCRIBE THEM. THE KEY CHALLENGE ADDRESSED IS THAT OF DETERMINING WHICH INFORMATION ONE NEEDS TO KEEP, TOGETHER WITH ONE’S DIGITAL ASSETS, SO THAT THEY CAN BE UNDERSTOOD AND USED IN THE LONG-TERM – IN OTHER WORDS, EXACTLY WHICH METADATA ONE NEEDS. THE BOOK WILL GREATLY BENEFIT BEGINNERS AND CURRENT PRACTITIONERS ALIKE. IT IS EQUALLY TARGETED AT DIGITAL PRESERVATION REPOSITORY MANAGERS AND METADATA ANALYSTS WHO ARE RESPONSIBLE FOR DIGITAL PRESERVATION METADATA, AS IT IS AT STUDENTS IN LIBRARY, INFORMATION AND ARCHIVAL SCIENCE DEGREE PROGRAMS OR RELATED FIELDS. FURTHER, IT CAN BE USED AT THE CONCEPTION STAGE OF A DIGITAL PRESERVATION SYSTEM OR FOR SELF-AUDITING AN EXISTING SYSTEM.

## OPEN SOURCE INTELLIGENCE TOOLS AND RESOURCES HANDBOOK - I-INTELLIGENCE 2019-08-17

2018 VERSION OF THE OSINT TOOLS AND RESOURCES HANDBOOK. THIS VERSION IS ALMOST THREE TIMES THE SIZE OF THE LAST PUBLIC RELEASE IN 2016. IT REFLECTS THE CHANGING INTELLIGENCE NEEDS OF OUR CLIENTS IN BOTH THE PUBLIC AND PRIVATE SECTOR, AS WELL AS THE MANY AREAS WE HAVE BEEN ACTIVE IN OVER THE PAST TWO YEARS.

## MAKING DATA SMARTER WITH IBM SPECTRUM DISCOVER: PRACTICAL AI SOLUTIONS - IVAYLO B. BOZHINOV 2020-10-19

MORE THAN 80% OF ALL DATA THAT IS COLLECTED BY ORGANIZATIONS IS NOT IN A STANDARD RELATIONAL DATABASE. INSTEAD, IT IS TRAPPED IN UNSTRUCTURED DOCUMENTS, SOCIAL MEDIA POSTS, MACHINE LOGS, AND SO ON. MANY ORGANIZATIONS FACE SIGNIFICANT CHALLENGES TO MANAGE THIS DELUGE OF UNSTRUCTURED DATA, SUCH AS THE FOLLOWING EXAMPLES: PINPOINTING AND ACTIVATING RELEVANT DATA FOR LARGE-SCALE ANALYTICS LACKING THE FINE-GRAINED VISIBILITY THAT IS NEEDED TO MAP DATA TO BUSINESS PRIORITIES REMOVING REDUNDANT, OBSOLETE, AND TRIVIAL (ROT) DATA IDENTIFYING AND CLASSIFYING SENSITIVE DATA IBM® SPECTRUM DISCOVER IS A MODERN METADATA MANAGEMENT SOFTWARE THAT PROVIDES DATA INSIGHT FOR PETABYTE-SCALE FILE AND

OBJECT STORAGE, STORAGE ON-PREMISES, AND IN THE CLOUD. THIS SOFTWARE ENABLES ORGANIZATIONS TO MAKE BETTER BUSINESS DECISIONS AND GAIN AND MAINTAIN A COMPETITIVE ADVANTAGE. IBM SPECTRUM® DISCOVER PROVIDES A RICH METADATA LAYER THAT ENABLES STORAGE ADMINISTRATORS, DATA STEWARDS, AND DATA SCIENTISTS TO EFFICIENTLY MANAGE, CLASSIFY, AND GAIN INSIGHTS FROM MASSIVE AMOUNTS OF UNSTRUCTURED DATA. IT IMPROVES STORAGE ECONOMICS, HELPS MITIGATE RISK, AND ACCELERATES LARGE-SCALE ANALYTICS TO CREATE COMPETITIVE ADVANTAGE AND SPEED CRITICAL RESEARCH. THIS IBM REDBOOKS® PUBLICATION PRESENTS SEVERAL USE CASES THAT ARE FOCUSED ON ARTIFICIAL INTELLIGENCE (AI) SOLUTIONS WITH IBM SPECTRUM DISCOVER. THIS BOOK HELPS STORAGE ADMINISTRATORS AND TECHNICAL SPECIALISTS PLAN AND IMPLEMENT AI SOLUTIONS BY USING IBM SPECTRUM DISCOVER AND SEVERAL OTHER IBM STORAGE PRODUCTS.

**LUCENE IN ACTION** - OTIS GOSPODNETIC 2010-07-08

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.

**APACHE SOLR 3.1 COOKBOOK** - RAFAŁ KUJALA 2011

THIS BOOK IS PART OF PACKT'S COOKBOOK SERIES; EACH CHAPTER LOOKS AT A DIFFERENT ASPECT OF WORKING WITH APACHE SOLR. THE RECIPES DEAL WITH COMMON PROBLEMS OF WORKING WITH SOLR BY USING EASY-TO-UNDERSTAND, REAL-LIFE EXAMPLES. THE BOOK IS NOT IN ANY WAY A COMPLETE APACHE SOLR REFERENCE AND YOU SHOULD SEE IT AS A HELPING HAND WHEN THINGS GET ROUGH ON YOUR JOURNEY WITH APACHE SOLR. DEVELOPERS WHO ARE WORKING WITH APACHE SOLR AND WOULD LIKE TO KNOW HOW TO COMBAT COMMON PROBLEMS WILL FIND THIS BOOK OF GREAT USE. KNOWLEDGE OF APACHE LUCENE WOULD BE A BONUS BUT IS NOT REQUIRED.

**TEXT PROCESSING WITH RUBY** - ROB MILLER 2015

“WHATEVER YOU WANT TO DO WITH TEXT, RUBY IS UP TO THE JOB. MOST INFORMATION IN THE WORLD IS IN TEXT FORMAT, AND YOU NEED TO MAKE SENSE OF THE DATA HIDING WITHIN. YOU WANT TO DO THIS EFFICIENTLY, AVOIDING LABOR-INTENSIVE, MANUAL WORK. TEXT PROCESSING WITH RUBY TAKES A PRACTICAL APPROACH TO WORKING WITH TEXT. FIRST, ACQUIRE: EXPLORE RUBY'S CORE AND STANDARD LIBRARY, AND EXTRACT TEXT INTO YOUR RUBY PROGRAMS. PROCESS DELIMITED FILES AND WEB PAGES, AND WRITE UTILITIES. SECOND, TRANSFORM: USE REGULAR EXPRESSIONS, WRITE A PARSER, AND USE NATURAL LANGUAGE PROCESSING TECHNIQUES. FINALLY, LOAD: WRITE THE TRANSFORMED TEXT AND DATA TO STANDARD OUTPUT, FILES, AND OTHER PROCESSES. SERIALIZE TEXT INTO JSON, XML, AND CVS, AND USE ERB TO CREATE MORE COMPLEX FORMATS. YOU'LL SOON BE ABLE TO TACKLE EVEN THE MOST ENORMOUS AND ENTANGLED TEXT WITH EASE.”--BACK COVER.

- SHUI YU 2016-03-03

THIS BOOK COVERS THREE MAJOR PARTS OF BIG DATA: CONCEPTS, THEORIES AND APPLICATIONS. WRITTEN BY WORLD-RENOWNED LEADERS IN BIG DATA, THIS BOOK EXPLORES THE PROBLEMS, POSSIBLE SOLUTIONS AND DIRECTIONS FOR BIG DATA IN RESEARCH AND PRACTICE. IT ALSO FOCUSES ON HIGH LEVEL CONCEPTS SUCH AS DEFINITIONS OF BIG DATA FROM DIFFERENT ANGLES; SURVEYS IN RESEARCH AND APPLICATIONS; AND EXISTING TOOLS, MECHANISMS, AND SYSTEMS IN PRACTICE. EACH CHAPTER IS INDEPENDENT FROM THE OTHER CHAPTERS, ALLOWING USERS TO READ ANY CHAPTER DIRECTLY. AFTER EXAMINING THE PRACTICAL SIDE OF BIG DATA, THIS BOOK PRESENTS THEORETICAL PERSPECTIVES. THE THEORETICAL RESEARCH RANGES FROM BIG DATA REPRESENTATION, MODELING AND TOPOLOGY TO DISTRIBUTION AND DIMENSION REDUCING. CHAPTERS ALSO INVESTIGATE THE MANY DISCIPLINES THAT INVOLVE BIG DATA, SUCH AS STATISTICS, DATA MINING, MACHINE LEARNING, NETWORKING, ALGORITHMS, SECURITY AND DIFFERENTIAL GEOMETRY. THE LAST SECTION OF THIS BOOK INTRODUCES BIG DATA APPLICATIONS FROM DIFFERENT COMMUNITIES, SUCH AS BUSINESS, ENGINEERING AND SCIENCE. BIG DATA CONCEPTS, THEORIES AND APPLICATIONS IS DESIGNED AS A REFERENCE FOR RESEARCHERS AND ADVANCED LEVEL STUDENTS IN COMPUTER SCIENCE, ELECTRICAL ENGINEERING AND MATHEMATICS. PRACTITIONERS WHO FOCUS ON INFORMATION SYSTEMS, BIG DATA, DATA MINING, BUSINESS ANALYSIS AND OTHER RELATED FIELDS WILL ALSO FIND THIS MATERIAL VALUABLE.

**ELASTICSEARCH COOKBOOK** - ALBERTO PARO 2013-12-24

WRITTEN IN AN ENGAGING, EASY-TO-FOLLOW STYLE, THE RECIPES WILL HELP YOU TO EXTEND THE CAPABILITIES OF ELASTICSEARCH TO MANAGE YOUR DATA EFFECTIVELY. IF YOU ARE A DEVELOPER WHO IMPLEMENTS ELASTICSEARCH IN YOUR WEB APPLICATIONS, MANAGE DATA, OR HAVE DECIDED TO START USING ELASTICSEARCH, THIS BOOK IS IDEAL FOR YOU. THIS BOOK ASSUMES THAT YOU'VE GOT WORKING KNOWLEDGE OF JSON AND JAVA

*BIG DATA CONCEPTS, THEORIES, AND APPLICATIONS*