Stanford Named Entity Recognition Ner Classifier

This book constitutes the thoroughly refereed post conference proceedings of the third edition of the Semantic Web Evaluation Challenge, SemWebEval 2016, co-located with the 13th European Semantic Web conference, held in Heraklion, Crete, Greece, in May/June 2016. This book includes the descriptions of all methods and tools that competed at SemWebEval 2016, together with a detailed description of the tasks, evaluation procedures and datasets. The contributions are grouped in the areas: Open Knowledge Extraction (OKE 2016); Semantic Sentiment Analysis (SSA 2016); Question Answering over Linked Data (QALD 6); Top-K Shortest Path in Large Typed RDF Graphs Datasets; Semantic Publishing (SemPub2016).

Research into computational models of argument is a rich interdisciplinary field involving the study of natural, artificial and theoretical argumentation and requiring openness to interactions with a variety of disciplines, ranging from philosophy and cognitive science to formal logic and graph theory. The ultimate aim is to support the development of computer-based systems able to engage in argumentation-related activities, either with human users or among themselves. This book presents the proceedings of the sixth biennial International Conference on Computational Models of Argument (COMMA 2016), held in Potsdam, Germany, on 12- 16 September. The aim of the COMMA conferences is to bring together researchers interested in computational models of argument and the representation of argumentation structures in natural language texts, with special attention to contributions concerning emerging trends and the development of new connections with other areas. The book contains the 25 full papers, 17 short papers and 10 demonstration abstracts presented at the conference, together with 3 invited talks. Subjects covered include abstract, bipolar and structured argumentation, quantitative approaches and their connections with formalisms like Bayesian networks and fuzzy logic, multi-agent scenarios, algorithms and solvers, and mining arguments in text, dialogue, and social media. The book provides an overview of current research and developments in the field of computational models of argument, and will be essential reading for all those with an interest in the field.

Get well-versed with traditional as well as modern natural language processing concepts and techniques Key Features Perform various NLP tasks to build linguistic applications using Python libraries Understand, analyze, and generate text to provide accurate results Interpret human language using various NLP concepts, methodologies, and tools Book Description Natural Language Processing (NLP) is the subfield in computational linguistics that enables computers to understand, process, and analyze text. This book caters to the unmet demand for hands-on training of NLP concepts and provides exposure to real-world applications along with a solid theoretical grounding. This book starts by introducing you to the field of NLP and its applications, along with the modern Python libraries that you'll use to build your NLP-powered apps. With the help of practical examples, you'll learn how to build reasonably sophisticated NLP applications, and cover various methodologies and challenges in deploying NLP applications in the real world. You'll cover key NLP tasks such as text classification, semantic embedding, sentiment analysis, machine translation, and developing a chatbot using machine learning and deep learning techniques. The book will also help you discover how machine learning techniques play a vital role in making your linguistic apps smart. Every chapter is accompanied by examples of real-world applications to help you build impressive NLP applications of your own. By the end of this NLP book, you'll be able to work with language data, use machine learning to identify patterns in text, and get acquainted with the advancements in NLP. What you will learn Understand how NLP powers modern applications Explore key NLP techniques to build your natural language vocabulary Transform text data into mathematical data structures and learn how to improve text mining models Discover how various neural network architectures work with natural language data Get the hang of building sophisticated text processing models using machine learning and deep learning Check out state-of-the-art architectures that have revolutionized research in the NLP domain Who this book is for This NLP Python book is for anyone looking to learn NLP's theoretical and practical aspects alike. It starts with the basics and gradually covers advanced concepts to make it easy to follow for readers with varying levels of NLP proficiency. This comprehensive guide will help you develop a thorough understanding of the NLP methodologies for building linguistic applications; however, working knowledge of Python programming language and high school level mathematics is expected.

This book provides a comprehensive introduction to the conversational interface, which is becoming the main mode of interaction with virtual personal assistants, smart devices, various types of wearable, and social robots. The book consists of four parts. Part I presents the background to conversational interfaces, examining past and present work on spoken language interaction with computers. Part II covers the various technologies that are required to build a conversational interface along with practical chapters and exercises using open source tools. Part III looks at interactions with smart devices, wearables, and robots, and discusses the role of emotion and personality in the conversational interface. Part IV examines methods for evaluating conversational interfaces and discusses future directions.

Leverage the power of Python to collect, process, and mine deep insights from social media data About This Book Acquire data from various social media platforms such as Facebook, Twitter, YouTube, GitHub, and more Analyze and extract actionable insights from your social data using various Python tools A highly practical guide to conducting efficient social media analytics at scale Who This Book Is For If you are a programmer or a data analyst familiar with the Python programming language and want to perform analyses of your social data to acquire valuable business insights, this book is for you. The book does not assume any prior knowledge of any data analysis tool or process. What You Will Learn Understand the basics of social media mining Use PyMongo to clean, store, and access data in MongoDB Understand user reactions and emotion detection on Facebook Perform Twitter sentiment analysis and entity recognition using Python Analyze video and campaign performance on YouTube Mine popular trends on GitHub and predict the next big technology Extract conversational topics on public internet forums Analyze user interests on Pinterest Perform large-scale social media analytics on the cloud In Detail Social Media platforms such as Facebook, Twitter, Forums, Pinterest, and YouTube have become part of everyday life in a big way. However, these complex and noisy data streams pose a potent challenge to everyone when it comes to harnessing them properly and benefiting from them. This book will introduce you to the concept of social media analytics, and how you can leverage its capabilities to empower your business. Right from acquiring data from various social networking sources such as Twitter, Facebook, YouTube, Pinterest, and social forums, you will see how to clean data and make it ready for analytical operations using various Python APIs. This book explains how to structure the clean data obtained and store in MongoDB using PyMongo. You will also perform web scraping and visualize data using Scrappy and

Beautifulsoup. Finally, you will be introduced to different techniques to perform analytics at scale for your social data on the cloud, using Python and Spark. By the end of this book, you will be able to utilize the power of Python to gain valuable insights from social media data and use them to enhance your business processes. Style and approach This book follows a step-by-step approach to teach readers the concepts of social media analytics using the Python programming language. To explain various data analysis processes, real-world datasets are used wherever required.

Many books and courses tackle natural language processing (NLP) problems with toy use cases and well-defined datasets. But if you want to build, iterate, and scale NLP systems in a business setting and tailor them for particular industry verticals, this is your guide. Software engineers and data scientists will learn how to navigate the maze of options available at each step of the journey. Through the course of the book, authors Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta, and Harshit Surana will guide you through the process of building real-world NLP solutions embedded in larger product setups. You'll learn how to adapt your solutions for different industry verticals such as healthcare, social media, and retail. With this book, you'll: Understand the wide spectrum of problem statements, tasks, and solution approaches within NLP Implement and evaluate different NLP applications using machine learning and deep learning methods Fine-tune your NLP solution based on your business problem and industry vertical Evaluate various algorithms and approaches for NLP product tasks, datasets, and stages Produce software solutions following best practices around release, deployment, and DevOps for NLP systems Understand best practices, opportunities, and the roadmap for NLP from a business and product leader's perspective

This three volume set (CCIS 853-855) constitutes the proceedings of the 17th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU 2017, held in Cádiz, Spain, in June 2018. The 193 revised full papers were carefully reviewed and selected from 383 submissions. The papers are organized in topical sections on advances on explainable artificial intelligence; aggregation operators, fuzzy metrics and applications; belief function theory and its applications; current techniques to model, process and describe time series; discrete models and computational intelligence; formal concept analysis and uncertainty; fuzzy implication functions; fuzzy logic and artificial intelligence problems; fuzzy mathematical analysis and applications; fuzzy methods in data mining and knowledge discovery; fuzzy transforms: theory and applications to data analysis and image processing; imprecise probabilities: foundations and applications; mathematical fuzzy logic, mathematical morphology; measures of comparison and entropies for fuzzy sets and their extensions; new trends in data aggregation; pre-aggregation functions and generalized forms of monotonicity; rough and fuzzy similarity modelling tools; soft computing for decision making in uncertainty; soft computing in information retrieval and sentiment analysis; tri-partitions and uncertainty; decision making modeling and applications; logical methods in mining knowledge from big data; metaheuristics and machine learning; optimization models for modern analytics; uncertainty in medicine; uncertainty in Video/Image Processing (UVIP).

This book addresses the impacts of various types of services such as infrastructure, platforms, software, and business processes that cloud computing and Big Data have introduced into business. Featuring chapters which discuss effective and efficient approaches in dealing with the inherent complexity and increasing demands in data science, a variety of application domains are covered. Various case studies by data management and analysis experts are presented in these chapters. Covered applications include banking, social networks, bioinformatics, healthcare, transportation and criminology. Highlighting the Importance of Big Data Management and Analysis for Various Applications will provide the reader with an understanding of how data management and analysis are adapted to these applications. This book will appeal to researchers and professionals in the field.

Having the ability to measure and explore the geographic space that surrounds us provides endless opportunities for us to utilize and interact with the world. As a broad field of study, geospatial research has applications in a variety of fields including military science, environmental science, civil engineering, and space exploration. Geospatial Research: Concepts, Methodologies, Tools, and Applications is a multi-volume publication highlighting critical topics related to geospatial analysis, geographic information systems, and geospatial technologies. Exploring multidisciplinary applications of geographic information systems and technologies in addition to the latest trends and developments in the field, this publication is ideal for academic and government library inclusion, as well as for reference by data scientists, engineers, government agencies, researchers, and graduate-level students in GIS programs.

This book constitutes the refereed post-conference proceedings of a Workshop focussing on Text Processing, held at the Forum for Information Retrieval Evaluation, FIRE 2016, in Kolkata, India, in December 2016. 16 full papers have been selected for inclusion in the book out of 19 submissions. The papers refer to the following seven tracks: Consumer Health Information Search (CHIS), Detecting Paraphrases in Indian Languages (DPIL), Information Extraction from Microblogs Posted during Disasters, Persian Plagiarism Detection (PersianPlagDet), Personality Recognition in SOurce COde (PR-SOCO), Shared Task on Mixed Script Information Retrieval (MSIR), and Shared Task on Code Mix Entity Extraction in Indian Languages (CMEE-IL).

Work with Python and powerful open source tools such as Gensim and spaCy to perform modern text analysis, natural language processing, and computational linguistics algorithms. Key Features Discover the open source Python text analysis ecosystem, using spaCy, Gensim, scikit-learn, and Keras Hands-on text analysis with Python, featuring natural language processing and computational linguistics algorithms Learn deep learning techniques for text analysis Book Description Modern text analysis is now very accessible using Python and open source tools, so discover how you can now perform modern text analysis in this era of textual data. This book shows you how to use natural language processing, and computational linguistics algorithms, to make inferences and gain insights about data you have. These algorithms are based on statistical machine learning and artificial intelligence techniques. The tools to work with these algorithms are available to you right now - with Python, and tools like Gensim and spaCy. You'll start by

learning about data cleaning, and then how to perform computational linguistics from first concepts. You're then ready to explore the more sophisticated areas of statistical NLP and deep learning using Python, with realistic language and text samples. You'll learn to tag, parse, and model text using the best tools. You'll gain hands-on knowledge of the best frameworks to use, and you'll know when to choose a tool like Gensim for topic models, and when to work with Keras for deep learning. This book balances theory and practical hands-on examples, so you can learn about and conduct your own natural language processing projects and computational linguistics. You'll discover the rich ecosystem of Python tools you have available to conduct NLP - and enter the interesting world of modern text analysis. What you will learn Why text analysis is important in our modern age Understand NLP terminology and get to know the Python tools and datasets Learn how to pre-process and clean textual data Convert textual data into vector space representations Using spaCy to process text Train your own NLP models for computational linguistics Use statistical learning and Topic Modeling algorithms for text, using Gensim and scikit-learn Employ deep learning techniques for text analysis using Keras Who this book is for This book is for you if you want to dive in, hands-first, into the interesting world of text analysis and NLP, and you're ready to work with the rich Python ecosystem of tools and datasets waiting for you!

This volume of the series "Translation and Multilingual Natural Language Processing" includes most of the papers presented at the Workshop "Language Technology for a Multilingual Europe", held at the University of Hamburg on September 27, 2011 in the framework of the conference GSCL 2011 with the topic "Multilingual Resources and Multilingual Applications", along with several additional contributions. In addition to an overview article on Machine Translation and two contributions on the European initiatives META-NET and Multilingual Web, the volume includes six full research articles. Our intention with this workshop was to bring together various groups concerned with the umbrella topics of multilingualism and language technology, especially multilingual technologies. This encompassed, on the one hand, representatives from research and development in the field of language technologies, and, on the other hand, users from diverse areas such as, among others, industry, administration and funding agencies. The Workshop "Language Technology for a Multilingual Europe" was co-organised by the two GSCL working groups "Text Technology" and "Machine Translation" (http://gscl.info) as well as by META-NET (http://www.meta-net.eu).

With the rise in data science development, we now have many remarkable techniques and tools to extend data analysis from numeric and categorical data to textual data. Sifting through the open-ended responses from a survey, for example, was an arduous process when performed by hand. Using a case study approach, this book was written for business analysts who wish to increase their skills in extracting answers for text data in order to support business decision making. Most of the exercises use Excel, today's most common analysis tool, and R, a popular analytic computer environment. The techniques covered range from the most basic text analytics, such as key word analysis, to more sophisticated techniques, such as topic extraction and text similarity scoring. Companion files with numerous datasets are included for use with case studies and exercises. FEATURES: Organized by tool or technique, with the basic techniques presented first and the more sophisticated techniques presented later Uses Excel and R for datasets in case studies and exercises Features the CRISP-DM data mining standard with early chapters for conducting the preparatory steps in data mining Companion files with numerous datasets and figures from the text.

Leverage Natural Language Processing (NLP) in Python and learn how to set up your own robust environment for performing text analytics. This second edition has gone through a major revamp and introduces several significant changes and new topics based on the recent trends in NLP. You'll see how to use the latest state-of-the-art frameworks in NLP, coupled with machine learning and deep learning models for supervised sentiment analysis powered by Python to solve actual case studies. Start by reviewing Python for NLP fundamentals on strings and text data and move on to engineering representation methods for text data, including both traditional statistical models and newer deep learning-based embedding models. Improved techniques and new methods around parsing and processing text are discussed as well. Text summarization and topic models have been overhauled so the book showcases how to build, tune, and interpret topic models in the context of an interest dataset on NIPS conference papers. Additionally, the book covers text similarity techniques with a real-world example of movie recommenders, along with sentiment analysis using supervised and unsupervised techniques. There is also a chapter dedicated to semantic analysis where you'll see how to build your own named entity recognition (NER) system from scratch. While the overall structure of the book remains the same, the entire code base, modules, and chapters has been updated to the latest Python 3.x release. What You'll Learn • Understand NLP and text syntax, semantics and structure• Discover text cleaning and feature engineering• Review text classification and text clustering • Assess text summarization and topic models• Study deep learning for NLP Who This Book Is For IT professionals, data analysts, developers, linguistic experts, data scientists and engineers and basically anyone with a keen interest in linguistics, analytics and generating insights from textual data.

This book constitutes the thoroughly refereed post-conference proceedings of the Second International Workshop on Algorithmic Aspects of Cloud Computing, ALGOCLOUD 2016, held in Aarhus, Denmark, in August 2016. The 11 revised full papers presented together with one tutorial paper were carefully reviewed and selected from 30 initial submissions. They deal with the following topics: algorithmic aspects of elasticity and scalability for distributed, large-scale data stores (e.g. NoSQL and columnar databases); search and retrieval algorithms for cloud infrastructures; monitoring and analysis of elasticity for virtualized environments; NoSQL, schemaless data modeling, integration; caching and load-balancing; storage structures and indexing for cloud databases; new algorithmic aspects of parallel and distributed computing for cloud applications; scalable machine learning, analytics and data science; high availability, reliability, failover; transactional models and algorithms for cloud databases; query languages and processing

programming models; consistency, replication and partitioning CAP, data structures and algorithms for eventually consistent stores.

The time is right for this all-new survey of the library technology that's already transitioning from trend to everyday reality. As in the previous best-selling volume, Varnum and his contributors throw the spotlight on the systems, software, and approaches most crucial to the knowledge institutions of tomorrow. Inside, readers will find concise information and analysis on topics such as mobile technologies; privacy-protection technology tools; the Internet of Things (IoT); virtual reality; bots and automation; machine learning applications for libraries; libraries as digital humanities enablers; visualizations in discovery systems; linked open data; embeddedness and Learning Tools Interoperability (LTI); special collections and digital publishing; link rot, web archiving, and the future of the Distributed Web; and digital repositories. Sure to spark discussions about library innovation, this collection is a must have for staff interested in technology or involved with strategic planning.

This book constitutes the refereed proceedings of the 16th Industrial Conference on Advances in Data Mining, ICDM 2016, held in New York, NY, USA, in July 2016. The 33 revised full papers presented were carefully reviewed and selected from 100 submissions. The topics range from theoretical aspects of data mining to applications of data mining, such as in multimedia data, in marketing, in medicine, and in process control, industry, and society.

This book constitutes the proceedings of the 6th International Information Retrieval Facility Conference, IRFC 2013, held in Limassol, Cyprus, October 2013. The 8 papers presented together with 2 short papers were carefully reviewed and selected from 16 high-quality submissions. IRF conferences wish to bring young researchers into contact with industry at an early stage. This sixth conference aimed to tackle four complementary research areas: information retrieval, machine translations for search solutions, and interactive information access.

This book provides practical information about web archives, offers inspiring examples for web archivists, raises new challenges, and shares recent research results about access methods to explore information from the past preserved by web archives. The book is structured in six parts. Part 1 advocates for the importance of web archives to preserve our collective memory in the digital era, demonstrates the problem of web ephemera and shows how web archiving activities have been trying to address this challenge. Part 2 then focuses on different strategies for selecting web content to be preserved and on the media types that different web archives host. It provides an overview of efforts to address the preservation of web content as well as smaller-scale but high-quality collections of social media or audiovisual content. Next, Part 3 presents examples of initiatives to improve access to archived web information and provides an overview of access mechanisms for web archives designed to be used by humans or automatically accessed by machines. Part 4 presents research use cases for web archives. It also discusses how to engage more researchers in exploiting web archives and provides inspiring research studies performed using the exploration of web archives. Subsequently, Part 5 demonstrates that web archives should become crucial infrastructures for modern connected societies. It makes the case for developing web archives as research infrastructures and presents several inspiring examples of added-value services built on web archives. Lastly, Part 6 reflects on the evolution of the web and the sustainability of web archiving activities. It debates the requirements and challenges for web archives if they are to assume the responsibility of being societal infrastructures that enable the preservation of memory. This book targets academics and advanced professionals in a broad range of research areas such as digital humanities, social sciences, history, media studies and information or computer science. It also aims to fill the need for a scholarly overview to support lecturers who would like to introduce web archiving into their courses by offering an initial reference for students.

This handbook covers a wide range of topics related to the collection, processing, analysis, and use of geospatial data in their various forms. This handbook provides an overview of how spatial computing technologies for big data can be organized and implemented to solve real-world problems. Diverse subdomains ranging from indoor mapping and navigation over trajectory computing to earth observation from space, are also present in this handbook. It combines fundamental contributions focusing on spatio-textual analysis, uncertain databases, and spatial statistics with application examples such as road network detection or colocation detection using GPUs. In summary, this handbook gives an essential introduction and overview of the rich field of spatial information science and big geospatial data. It introduces three different perspectives, which together define the field of big geospatial data: a societal, governmental, and governance perspective. It discusses questions of how the acquisition, distribution and exploitation of big geospatial data must be organized both on the scale of companies and countries. A second perspective is a theory-oriented set of contributions on arbitrary spatial data with contributions introducing into the exciting field of spatial statistics or into uncertain databases. A third perspective is taking a very practical perspective to big geospatial data, ranging from chapters that describe how big geospatial data infrastructures can be implemented and how specific applications can be implemented on top of big geospatial data. This would include for example, research in historic map data, road network extraction, damage estimation from remote sensing imagery, or the analysis of spatio-textual collections and social media. This multi-disciplinary approach makes the book unique. This handbook can be used as a reference for undergraduate students, graduate students and researchers focused on big geospatial data. Professionals can use this book, as well as practitioners facing big collections of geospatial data.

Design and Implementation of service-oriented architectures imposes a huge number of research questions from the fields of software engineering, system analysis and modeling, adaptability, and application integration. Component orientation and web services are two approaches for design and realization of complex web-based system. Both approaches allow for dynamic application adaptation as well as integration of enterprise application. Commonly used technologies, such as J2EE and .NET, form de facto standards for the realization of complex distributed systems. Evolution of component systems has lead to web services and service-based architectures. This has been

manifested in a multitude of industry standards and initiatives such as XML, WSDL UDDI, SOAP, etc. All these achievements lead to a new and promising paradigm in IT systems engineering which proposes to design complex software solutions as collaboration of contractually defined software services. Service-Oriented Systems Engineering represents a symbiosis of best practices in object-orientation, component-based development, distributed computing, and business process management. It provides integration of business and IT concerns. The annual Ph.D. Retreat of the Research School provides each member the opportunity to present his/her current state of their research and to give an outline of a prospective Ph.D. thesis. Due to the interdisciplinary structure of the research school, this technical report covers a wide range of topics. These include but are not limited to: Human Computer Interaction and Computer Vision as Service; Service-oriented Geovisualization Systems; Algorithm Engineering for Service-oriented Systems; Modeling and Verification of Self-adaptive Service-oriented Systems; Tools and Methods for Software Engineering in Service-oriented Systems; Security Engineering of Service-based IT Systems; Service-oriented Information Systems; Evolutionary Transition of Enterprise Applications to Service Orientation; Operating System Abstractions for Service-oriented Computing; and Services Specification, Composition, and Enactment.

"Practical Data Analysis – Using Python & Open Source Technology" uses a case-study based approach to explore some of the real-world applications of open source data analysis tools and techniques. Specifically, the following topics are covered in this book: 1. Open Source Data Analysis Tools and Techniques. 2. A Beginner's Guide to "Python" for Data Analysis. 3. Implementing Custom Search Engines On The Fly. 4. Visualising Missing Data. 5. Sentiment Analysis and Named Entity Recognition. 6. Automatic Document Classification, Clustering and Summarisation. 7. Fraud Detection Using Machine Learning Techniques. 8. Forecasting - Using Data to Map the Future. 9. Continuous Monitoring and Real-Time Analytics. 10. Creating a Robot for Interacting with Web Applications. Free samples of the book is available at - http://timesofdatascience.com

This book constitutes the refereed proceedings of the 4th Annual International Symposium on Information Management and Big Data, SIMBig 2017, held in Lima, Peru, in September 2017. The 10 revised full papers presented were carefully reviewed and selected from 71 submissions. The papers address issues such as Data Science, Big Data, Data Mining, Natural Language Processing, Text Mining, Information Retrieval, Machine Learning, Semantic Web, Ontologies, Web Mining, Knowledge Representation and Linked Open Data, Social Web and Web Science, Information Visualization.

The two-volume set LNAI 7301 and 7302 constitutes the refereed proceedings of the 16th Pacific-Asia Conference on Knowledge Discovery and Data Mining, PAKDD 2012, held in Kuala Lumpur, Malaysia, in May 2012. The total of 20 revised full papers and 66 revised short papers were carefully reviewed and selected from 241 submissions. The papers present new ideas, original research results, and practical development experiences from all KDD-related areas. The papers are organized in topical sections on supervised learning: active, ensemble, rare-class and online; unsupervised learning: clustering, probabilistic modeling in the first volume and on pattern mining: networks, graphs, time-series and outlier detection, and data manipulation: pre-processing and dimension reduction in the second volume.

This book constitutes the refereed proceedings of the 15th International Semantic Web Conference, ESWC 2018, held in Heraklion, Crete, Greece. The 48 revised full papers presented were carefully reviewed and selected from 179 submissions. The papers cover a large range of topics such as logical modelling and reasoning, natural language processing, databases and data storage and access, machine learning, distributed systems, information retrieval and data mining, social networks, and Web science and Web engineering.

Linguistic annotation and text analytics are active areas of research and development, with academic conferences and industry events such as the Linguistic Annotation Workshops and the annual Text Analytics Summits. This book provides a basic introduction to both fields, and aims to show that good linguistic annotations are the essential foundation for good text analytics. After briefly reviewing the basics of XML, with practical exercises illustrating in-line and stand-off annotations, a chapter is devoted to explaining the different levels of linguistic annotations. The reader is encouraged to create example annotations using the WordFreak linguistic annotation tool. The next chapter shows how annotations can be created automatically using statistical NLP tools, and compares two sets of tools, the OpenNLP and Stanford NLP tools. The second half of the book describes different annotation formats and gives practical examples of how to interchange annotations between different 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. Copies of the example files, scripts, and stylesheets used in the book are available from the companion website, located at the book website. Table of Contents: Working with XML / Linguistic Annotation / Using Statistical NLP Tools / Annotation Interchange / Annotation Architectures / Text Analytics

This book constitutes the refereed proceedings of the 6th IAPR TC3 International Workshop on Artificial Neural Networks in Pattern Recognition, ANNPR 2014, held in Montreal, QC, Canada, in October 2014. The 24 revised full papers presented were carefully reviewed and selected from 37 submissions for inclusion in this volume. They cover a large range of topics in the field of learning algorithms and architectures and discussing the latest research, results, and ideas in these areas.

The two-volume set LNCS 10704 and 10705 constitutes the thoroughly refereed proceedings of the 24th International Conference on Multimedia Modeling, MMM 2018, held in Bangkok, Thailand, in February 2018. Of the 185 full papers submitted, 46 were selected for oral presentation and 28 for poster presentation; in addition, 5 papers were accepted for Multimedia Analytics: Perspectives, Techniques, and Applications, 12 extended abstracts for demonstrations ,and 9 accepted papers for Video Browser Showdown 2018. All papers presented were carefully reviewed and selected from 185 submissions.

Can established humanities methods coexist with computational thinking? It is one of the major questions in humanities research today, as scholars increasingly adopt sophisticated data science for their work. Though the study and interpretation of texts alongside sophisticated computational tools can serve scholarship, these methods cannot replace existing frameworks. As Dobson shows, ideas of scientific validity cannot easily nor should be adapted for humanities research because digital humanities, unlike science, lack a leading-edge horizon charting the frontiers of inquiry. Instead, the methods of digital humanities require a constant rereading. At the same time, suspicious and critical readings of digital methodologies make it unwise for scholars to defer to computational methods. Humanists must examine the tools--including the assumptions that went into the codes and algorithms--and questions surrounding their own use of digital technology in research. Insightful and forward thinking, Critical Digital Humanities lays out a new path of humanistic inquiry that merges critical theory and computational science.

Speech & Language ProcessingPearson Education IndiaIntroduction to Linguistic Annotation and Text AnalyticsMorgan & Claypool Publishers

Learn to build expert NLP and machine learning projects using NLTK and other Python libraries About This Book Break text down into its component parts for spelling correction, feature extraction, and

phrase transformation Work through NLP concepts with simple and easy-to-follow programming recipes Gain insights into the current and budding research topics of NLP Who This Book Is For If you are an NLP or machine learning enthusiast and an intermediate Python programmer who wants to quickly master NLTK for natural language processing, then this Learning Path will do you a lot of good. Students of linguistics and semantic/sentiment analysis professionals will find it invaluable. What You Will Learn The scope of natural language complexity and how they are processed by machines Clean and wrangle text using tokenization and chunking to help you process data better Tokenize text into sentences and sentences into words Classify text and perform sentiment analysis Implement string matching algorithms and normalization techniques Understand and implement the concepts of information retrieval and text summarization Find out how to implement various NLP tasks in Python In Detail Natural Language Processing is a field of computational linguistics and artificial intelligence that deals with human-computer interaction. It provides a seamless interaction between computers and human beings and gives computers the ability to understand human speech with the help of machine learning. The number of human-computer interaction instances are increasing so it's becoming imperative that computers comprehend all major natural languages. The first NLTK Essentials module is an introduction on how to build systems around NLP, with a focus on how to create a customized tokenizer and parser from scratch. You will learn essential concepts of NLP, be given practical insight into open source tool and libraries available in Python, shown how to analyze social media sites, and be given tools to deal with large scale text. This module also provides a workaround using some of the amazing capabilities of Python libraries such as NLTK, scikit-learn, pandas, and NumPy. The second Python 3 Text Processing with NLTK 3 Cookbook module teaches you the essential techniques of text and language processing with simple, straightforward examples. This includes organizing text corpora, creating your own custom corpus, text classification with a focus on sentiment analysis, and distributed text processing methods. The third Mastering Natural Language Processing with Python module will help you become an expert and assist you in creating your own NLP projects using NLTK. You will be guided through model development with machine learning tools, shown how to create training data, and given insight into the best practices for designing and building NLP-based applications using Python. This Learning Path combines some of the best that Packt has to offer in one complete, curated package and is designed to help you quickly learn text processing with Python and NLTK. It includes content from the following Packt products: NTLK essentials by Nitin Hardeniya Python 3 Text Processing with NLTK 3 Cookbook by Jacob Perkins Mastering Natural Language Processing with Python by Deepti Chopra, Nisheeth Joshi, and Iti Mathur Style and approach This comprehensive course creates a smooth learning path that teaches you how to get started with Natural Language Processing using Python and NLTK. You'll learn to create effective NLP and machine learning projects using Python and NLTK.

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.

With higher education turning towards data analytics as the next big advance in technology, it is important to look at how information is gathered and visualized for accurate comprehension, analysis, and decision-making. Packaging Digital Information for Enhanced Learning and Analysis: Data Visualization, Spatialization, and Multidimensionality brings together effective practices for the end-to-end capture and web based presentation of information for comprehension, analysis, and decision-making. This publication is beneficial for educators, trainers, instructional designers, web designers, and graduate students interested in improving analytical tools.

This book constitutes the proceedings of the Sixth Conference on Information and Communication Technologies "TIC.EC", held in Cuenca, Ecuador, from November 27 to 29, 2019. Considered one of the most important conferences on ICT in Ecuador, it brings together scholars and practitioners from the country and abroad to discuss the development, issues and projections of the use of information and communication technologies in multiples fields of application. The 2019 "TIC.EC" conference was organized by Universidad del Azuay (UDA) and its Engineering School, as well as the Ecuadorian Corporation for the Development of Research and Academia (CEDIA). The book covers the following topics: · Software engineering · Security · Data · Networks · Architecture · Applied ICTs · Technological entrepreneurship · Links between research and industry · High-impact innovation · Knowledge management and intellectual property

This book showcases various ways in which digital archives allow for new approaches to journalism history. The chapters in this book were selected based on three overall objectives: 1) research that highlights specific concerns within journalism history through digital archives; 2) discussions of digital methodologies, as well as specific applications, that are accessible for journalism scholars with no prior experiences with such approaches; and 3) that journalism history and digital archives are connected in other ways than through specific methods, i.e., that the connection raises larger questions of historiography and power. The contributions address cases and developments in Asia, South and North America and Europe; and range from long-range, big-data, machine-leaning and topic modelling studies of journalistic characteristics and meta-journalistic discourses to critiques of archival practices and access in relation to gender, social movements and poverty. The chapters in this book were originally published as a special issue of Digital Journalism.