Computational Linguistics An Introduction

In a globalized society, effective communication is critical, and study of language from a mathematical perspective can shed light on new ways in which to express meaning across cultures and nations. Computational Linguistics: Concepts, Methodologies, Tools, and Applications explores language by dissecting the phonemic aspects of various communication systems in order to identify similarities and pitfalls in the expression of meaning. With applications in a variety of areas, from psycholinguistics and cognitive science to computer science and artificial intelligence, this multivolume reference work will be of use to researchers, professionals, and educators on the cutting edge of language acquisition and communication science.

This book constitutes the refereed proceedings of the 10th International Conference on Computational Linguistics and Intelligent Text Processing, CICLing 2009, held in Mexico City, Mexico in March 2009. The 44 revised full papers presented together with 4 invited papers were carefully reviewed and selected from numerous submissions. The papers cover all current issues in computational linguistics research and present intelligent text processing applications.

Overview of the interface of language and the law, illustrated with authentic data and contemporary case studies. Topics include collection of evidence, discourse, courtroom interaction, legal language, comprehension and forensic phonetics.

Provides a clearly-written, concise and accessible introduction to speech and language processing, with accompanying software. Language and Computers introduces students to thefundamentals of how computers are used to represent, process, andorganize textual and spoken information. Concepts are grounded inreal-world examples familiar to students' experiences ofusing language and computers in everyday life. A real-world introduction to the fundamentals of how computersprocess language, written specifically for the undergraduateaudience, introducing key concepts from computationallinguistics. Offers a comprehensive explanation of the problems computersface in handling natural language Covers a broad spectrum of language-related applications andissues, including major computer applications involving naturallanguage and the social and ethical implications of these newdevelopments The book focuses on real-world examples with which students canidentify, using these to explore the technology and how itworks Features "under-the-hood" sections that givegreater detail on selected advanced topics, rendering the bookappropriate for more advanced courses, or for independent study bythe motivated reader.

This book focuses mainly on logical approaches to computational linguistics, but also discusses integrations with other approaches, presenting both classic and newly emerging theories and applications.Decades of research on theoretical work and practical applications have demonstrated that computational linguistics is a distinctively interdisciplinary area. There is convincing evidence that computational approaches to linguistics can benefit from research on the nature of human language, including from the perspective of its evolution. This book addresses various topics in computational theories of human language, covering grammar, syntax, and semantics. The common thread running through the research presented is the role of computer science, mathematical logic and other subjects of mathematics in computational linguistics and natural language processing (NLP).

Promoting intelligent approaches to artificial intelligence (AI) and NLP, the book is intended for researchers and graduate students in the field.

"Solving linguistic problems frequently reduces to carrying out tasks that are computationally complex and therefore requires automation. This book is an introduction to machine-aided linguistic discovery, a novel research area, and argues for the fruitfulness of the computational approach by presenting a basic conceptual apparatus and several intelligent discovery programs. One of the programs models the fundamental Saussurian notion of 'system' and thus, almost a century after the introduction of this concept and structuralism in general, linguists are for the first time capable of handling adequately this recurring computationally complex task. Another program models the problem of searching for Greenbergian language universals and is capable of stating its discoveries in an intelligible form, a comprehensive English language text. It is the first computer program to generate a whole scientific article. A third program detects potential inconsistencies in genetic language classifications. These, and the other programs described in this book, are applied with noteworthy results to substantial problems from diverse linguistic disciplines such as structural semantics, phonology, typology and historical linguistics."--Publisher's description.

The rapid advancement in the theoretical understanding of statistical and machine learning methods for semisupervised learning has made it difficult for nonspecialists to keep up to date in the field. Providing a broad, accessible treatment of the theory as well as linguistic applications, Semisupervised Learning for Computational Linguistics offers self-contained coverage of semisupervised methods that includes background material on supervised and unsupervised learning. The book presents a brief history of semisupervised learning and its place in the spectrum of learning methods before moving on to discuss well-known natural language processing methods, such as self-training and co-training. It then centers on machine learning techniques, including the boundary-oriented methods of perceptrons, boosting, support vector machines (SVMs), and the null-category noise model. In addition, the book covers clustering, the expectation-maximization (EM) algorithm, related generative methods, and agreement methods. It concludes with the graph-based method of label propagation as well as a detailed discussion of spectral methods. Taking an intuitive approach to the material, this lucid book facilitates the application of semisupervised learning methods to natural language processing and provides the framework and motivation for a more systematic study of machine learning.

A state-of-the-art reference to one of the most active and productive fields in linguistics: computational linguistics. Thirty-eight chapters, commissioned from experts all over the world, describe the major concepts, methods, and applications. Part I provides an overview of the field; Part II describes current tasks, techniques, and tools in natural language processing; and Part III surveys current applications.

How different are sign languages across the world? Are individual signs and signed sentences constructed in the same way across these languages? What are the rules for having a conversation in a sign language? How do children and adults learn a sign language? How are sign languages processed in the brain? These questions and many more are addressed in this introductory book on sign linguistics using examples from more than thirty different sign languages. Comparisons are also made with spoken languages. This book can be used as a self-study book or as a text book for students of sign linguistics. Each chapter concludes with a summary, some test-yourself questions and assignments, as well as a list of recommended texts for further reading. The book is accompanied by a website containing assignments, video clips and links to web resources.

A highly respected introduction to the computer analysis of language. Copyright © Libri GmbH. All rights reserved.

This comprehensive reference work provides an overview of the concepts, methodologies, and applications in computational linguistics and natural language processing (NLP). Features contributions by the top researchers in the field, reflecting the work that is driving the discipline forward Includes an introduction to the major theoretical issues in these fields, as well as the central engineering applications that the work has produced Presents the major developments in an accessible way, explaining the close connection between scientific understanding of the computational properties of natural language and the creation of effective language technologies Serves as an invaluable state-of-the-art reference source for computational linguists and software engineers developing NLP applications in industrial research and development labs of software companies

This book presents a collection of original research articles that showcase the state of the art of research in corpus and computational linguistic approaches to Chinese language teaching, learning and assessment. It offers a comprehensive set of corpus resources and natural language processing tools that are useful for teaching, learning and assessing Chinese as a second or foreign language; methods for implementing such resources and techniques in Chinese pedagogy and assessment; as well as research findings on the effectiveness of using such resources and techniques in various aspects of Chinese pedagogy and assessment.

This accessible textbook is the only introduction to linguistics in which each chapter is written by an expert who teaches courses on that topic, ensuring balanced and uniformly excellent coverage of the full range of modern linguistics. Assuming no prior knowledge the text offers a clear introduction to the traditional topics of structural linguistics (theories of sound, form, meaning, and language change), and in addition provides full coverage of contextual linguistics, including separate chapters on discourse, dialect variation, language and culture, and the politics of language. There are also up-to-date separate chapters on language and the brain, computational linguistics, writing, child language acquisition, and second-language learning. The breadth of the textbook makes it ideal for introductory courses on language and linguistics offered by departments of English, sociology, anthropology, and communications, as well as by linguistics departments.

In this book, Almerindo E. Ojeda offers a unique perspective on linguistics by discussing developing computer programs that will assign particular sounds to particular meanings and, conversely, particular meanings to particular sounds. Since these assignments are to operate efficiently over unbounded domains of sound and sense, they can begin to model the two fundamental modalities of human language—speaking and hearing. The computational approach adopted in this book is motivated by our struggle with one of the key problems of contemporary linguistics—figuring out how it is that language emerges from the brain.

Computational LinguisticsAn IntroductionCambridge University Press

Statistics for Linguists: An Introduction Using R is the first statistics textbook on linear models for linguistics. The book covers simple uses of linear models through generalized models to more advanced approaches, maintaining its focus on conceptual issues and avoiding excessive mathematical details. It contains many applied examples using the R statistical programming environment. Written in an accessible tone and style, this text is the ideal main resource for graduate and advanced undergraduate students of Linguistics statistics courses as well as those in other fields, including Psychology, Cognitive Science, and Data Science.

"Focusing on the descriptive facts of English, this volume provides a systematic introduction to English syntax for students with no prior knowledge of English grammar or syntactic analysis. English Syntax aims to help students appreciate the various sentence patterns available in the language, understand insights into core data of its syntax, develop analytic abilities to further explore the patterns of English, and learn

precise ways of formalizing syntactic analysis for a variety of English data and major constructions such as agreement, raising and control, the auxiliary system, passive, wh- questions, relative clauses, extrapolation, and clefts"--Publisher's description.

This book provides system developers and researchers in natural language processing and computational linguistics with the necessary background information for working with the Arabic language. The goal is to introduce Arabic linguistic phenomena and review the state-of-the-art in Arabic processing. The book discusses Arabic script, phonology, orthography, morphology, syntax and semantics, with a final chapter on machine translation issues. The chapter sizes correspond more or less to what is linguistically distinctive about Arabic, with morphology getting the lion's share, followed by Arabic script. No previous knowledge of Arabic is needed. This book is designed for computer scientists and linguists alike. The focus of the book is on Modern Standard Arabic; however, notes on practical issues related to Arabic dialects and languages written in the Arabic script are presented in different chapters. Table of Contents: What is "Arabic"? / Arabic Script / Arabic Phonology and Orthography / Arabic Morphology / Computational Morphology Tasks / Arabic Syntax / A Note on Arabic Semantics / A Note on Arabic and Machine Translation

The latest edition of a popular introductory linguistics text, now including a section on computational linguistics, new non-English examples, quizzes for each chapter, and additional special topics. This popular introductory linguistics text is unique for its integration of themes. Rather than treat morphology, phonetics, phonology, syntax, and semantics as completely separate fields, the book shows how they interact. The authors provide a sound introduction to linguistic methodology, focusing on a set of linguistic concepts that are among the most fundamental within the field. By studying the topics in detail, students can get a feeling for how work in different areas of linguistics is done. As in the last edition, part I covers the structural and interpretive parts of language—morphology, phonetics, phonology, syntax, semantics, variation, and change. Part II covers use and context of language and includes chapters on pragmatics, psychology of language, language acquisition, and language and the brain. This seventh edition has been extensively revised and updated; new material includes a chapter on computational linguistics (available in digital form and updated regularly to reflect the latest research in a rapidly developing field), more non-English examples, and a wide range of exercises, quizzes, and special topics. The seventh edition of Linguistics includes access to a new, web-based eCourse and enhanced eTextbook. The content from the former print supplement A Linguistics Workbook is now available in this online eCourse as interactive exercises. The eCourse is available via the Rent eTextbook link at http://mitpress.mit.edu/linguistics7, and may be used on its own for self-study or integrated with instructor-led learning management systems. The eCourse is a comprehensive, web-based eLearning solution. There is nothing to download or install; it is accessible through any modern web browser and most mobile devices. It features a singular new tool for building syntax trees, an IPA keyboard, a combination of auto-graded and essay questions, and classroom management tools. The enhanced eTextbook includes videos and flashcards and allows bookmarking, note-taking, highlighting, and annotation sharing. Access to the eCourse is free with the purchase of a new textbook or e-book. New print copies of this book include a card affixed to the inside back cover with a unique access code for the eTextbook. If you purchased an e-book, you may obtain a unique access code by emailing digitalproducts-cs@mit.edu or calling 617-253-2889 or 800-207-8354 (toll-free in the U.S. and Canada). If you have a used copy of this book, you may purchase a digitally delivered access code separately via the Rent eTextbook link at http://mitpress.mit.edu/linguistics7.

Biomedical Natural Language Processing is a comprehensive tour through the classic and current work in the field. It discusses all subjects from both a rule-based and a machine learning approach, and also describes each subject from the perspective of both biological science

and clinical medicine. The intended audience is readers who already have a background in natural language processing, but a clear introduction makes it accessible to readers from the fields of bioinformatics and computational biology, as well. The book is suitable as a reference, as well as a text for advanced courses in biomedical natural language processing and text mining.

How can computers distinguish the coherent from the unintelligible, recognize new information in a sentence, or draw inferences from a natural language passage? Computational semantics is an exciting new field that seeks answers to these questions, and this volume is the first textbook wholly devoted to this growing subdiscipline. The book explains the underlying theoretical issues and fundamental techniques for computing semantic representations for fragments of natural language. This volume will be an essential text for computer scientists, linguists, and anyone interested in the development of computational semantics.

This is the first volume of a unique collection that brings together the best English-language problems created for students competing in the Computational Linguistics Olympiad. These problems are representative of the diverse areas presented in the competition and designed with three principles in mind: · To challenge the student analytically, without requiring any explicit knowledge or experience in linguistics or computer science; · To expose the student to the different kinds of reasoning required when encountering a new phenomenon in a language, both as a theoretical topic and as an applied problem; · To foster the natural curiosity students have about the workings of their own language, as well as to introduce them to the beauty and structure of other languages; · To learn about the models and techniques used by computers to understand human language. Aside from being a fun intellectual challenge, the Olympiad mimics the skills used by researchers and scholars in the field of computational linguistics. In an increasingly global economy where businesses operate across borders and languages, having a strong pool of computational linguists is a competitive advantage, and an important component to both security and growth in the 21st century. This collection of problems is a wonderful general introduction to the field of linguistics through the analytic problem solving technique. "A fantastic collection of problems for anyone who is curious about how human language works! These books take serious scientific questions and present them in a fun, accessible way. Readers exercise their logical thinking capabilities while learning about a wide range of human languages, linguistic phenomena, and computational models. " - Kevin Knight, USC Information Sciences Institute

An introduction to Python programming for linguists. Examples of code specifically designed for language analysis are featured throughout.

This book encompasses a collection of topics covering recent advances that are important to the Arabic language in areas of natural language processing, speech and image analysis. This book presents state-of-the-art reviews and fundamentals as well as applications and recent innovations.The book chapters by top researchers present basic concepts and challenges for the Arabic language in linguistic processing, handwritten recognition, document analysis, text classification and speech processing. In addition, it reports on selected applications in sentiment analysis, annotation, text summarization, speech and font analysis, word recognition and spotting and question answering.Moreover, it highlights and introduces some novel applications in vital areas for the Arabic language. The book is therefore a useful resource for young researchers who are interested in the Arabic language and are still developing their fundamentals and skills in this area. It is also interesting for scientists who wish to keep track of the most recent research directions and advances in this area.

The latest edition of a popular introductory text, including a section on computational linguistics, new non-English examples, quizzes for each chapter, and additional special topics. With this edition, Linguistics and content formerly included in A Linguistics Workbook are available as an interactive online textbook. Access codes are required for the interactive eTextbook. New print copies of this book include a card affixed to

the inside back cover with a unique access code. If you purchased a used copy of this book, this is a separately purchased printed access card.

Corpus Linguistics: An Introduction will appeal to a wide spectrum of scholars, researchers, and particularly to students of linguistics. It offers guidelines for the creation and usage of corpora in the form of empirical language databases with direct functional and theoretical interpretation of a natural language. Drawn from original research and written in an accessible language and style, this book will create avenues for further advancements in mainstream and applied linguistics and language technology.

The use of large, computerized bodies of text for linguistic analysis and description has emerged in recent years as one of the most significant and rapidly-developing fields of activity in the study of language. This book provides a comprehensive introduction and guide to Corpus Linguistics. All aspects of the field are explored, from the various types of electronic corpora that are available to instructions on how to design and compile a corpus. Graeme Kennedy surveys the development of corpora for use in linguistic research, looking back to the pre-electronic age as well as to the massive growth of computer corpora in the electronic age.

Weighted finite-state transducers (WFSTs) are commonly used by engineers and computational linguists for processing and generating speech and text. This book first provides a detailed introduction to this formalism. It then introduces Pynini, a Python library for compiling finite-state grammars and for combining, optimizing, applying, and searching finite-state transducers. This book illustrates this library's conventions and use with a series of case studies. These include the compilation and application of context-dependent rewrite rules, the construction of morphological analyzers and generators, and text generation and processing applications.

In recent years, online social networking has revolutionized interpersonal communication. The newer research on language analysis in social media has been increasingly focusing on the latter's impact on our daily lives, both on a personal and a professional level. Natural language processing (NLP) is one of the most promising avenues for social media data processing. It is a scientific challenge to develop powerful methods and algorithms which extract relevant information from a large volume of data coming from multiple sources and languages in various formats or in free form. We discuss the challenges in analyzing social media texts in contrast with traditional documents. Research methods in information extraction, automatic categorization and clustering, automatic summarization and indexing, and statistical machine translation need to be adapted to a new kind of data. This book reviews the current research on NLP tools and methods for processing the non-traditional information from social media data that is available in large amounts (big data), and shows how innovative NLP approaches can integrate appropriate linguistic information in various fields such as social media monitoring, healthcare, business intelligence, industry, marketing, and security and defence. We review the existing evaluation metrics for NLP and social media applications, and the new efforts in evaluation campaigns or shared tasks on new datasets collected from social media. Such tasks are organized by the Association for Computational Linguistics (such as SemEval tasks) or by the National Institute of Standards and Technology via the Text REtrieval Conference (TREC) and the Text Analysis Conference (TAC). In the concluding chapter, we discuss the importance of this dynamic discipline and its great potential for NLP in the coming decade, in the context of changes in mobile technology, cloud

computing, virtual reality, and social networking. In this second edition, we have added information about recent progress in the tasks and applications presented in the first edition. We discuss new methods and their results. The number of research projects and publications that use social media data is constantly increasing due to continuously growing amounts of social media data and the need to automatically process them. We have added 85 new references to the more than 300 references from the first edition. Besides updating each section, we have added a new application (digital marketing) to the section on media monitoring and we have augmented the section on healthcare applications with an extended discussion of recent research on detecting signs of mental illness from social media.

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!

The central task of future-oriented computational linguistics is the development of cognitive machines which humans can freely speak to in their natural language. This will involve the development of a functional theory of language, an objective method of verification, and a wide range of practical applications. Natural communication requires not only verbal processing, but also non-

verbal perception and action. Therefore, the content of this book is organized as a theory of language for the construction of talking robots with a focus on the mechanics of natural language communication in both the listener and the speaker.

This is the second volume of a unique collection that brings together the best English-language problems created for students competing in the Computational Linguistics Olympiad. These problems are representative of the diverse areas presented in the competition and designed with three principles in mind: · To challenge the student analytically, without requiring any explicit knowledge or experience in linguistics or computer science; · To expose the student to the different kinds of reasoning required when encountering a new phenomenon in a language, both as a theoretical topic and as an applied problem; · To foster the natural curiosity students have about the workings of their own language, as well as to introduce them to the beauty and structure of other languages; · To learn about the models and techniques used by computers to understand human language. Aside from being a fun intellectual challenge, the Olympiad mimics the skills used by researchers and scholars in the field of computational linguistics. In an increasingly global economy where businesses operate across borders and languages, having a strong pool of computational linguists is a competitive advantage, and an important component to both security and growth in the 21st century. This collection of problems is a wonderful general introduction to the field of linguistics through the analytic problem solving technique. "A fantastic collection of problems for anyone who is curious about how human language works! These books take serious scientific questions and present them in a fun, accessible way. Readers exercise their logical thinking capabilities while learning about a wide range of human languages, linguistic phenomena, and computational models. " - Kevin Knight, USC Information Sciences Institute

A survey of computational methods for understanding, generating, and manipulating human language, which offers a synthesis of classical representations and algorithms with contemporary machine learning techniques. This textbook provides a technical perspective on natural language processing—methods for building computer software that understands, generates, and manipulates human language. It emphasizes contemporary data-driven approaches, focusing on techniques from supervised and unsupervised machine learning. The first section establishes a foundation in machine learning by building a set of tools that will be used throughout the book and applying them to word-based textual analysis. The second section introduces structured representations of language, including sequences, trees, and graphs. The third section explores different approaches to the representation and analysis of linguistic meaning, ranging from formal logic to neural word embeddings. The final section offers chapter-length treatments of three transformative applications of natural language processing: information extraction, machine translation, and text generation. End-of-chapter exercises include both paper-and-pencil analysis and software implementation. The text synthesizes and distills a broad and diverse research literature, linking contemporary machine learning techniques with the field's linguistic and computational foundations. It is suitable for use in advanced undergraduate and graduate-level courses and as a reference for software engineers and data scientists. Readers should have a background in computer programming and college-level mathematics. After mastering the material presented, students will have the technical skill to build and analyze novel

natural language processing systems and to understand the latest research in the field.