

## <<Python自然语言处理>>

### 图书基本信息

书名：<<Python自然语言处理>>

13位ISBN编号：9787564122614

10位ISBN编号：7564122617

出版时间：2010-6

出版时间：东南大学出版社

作者：(英)伯德,(英)克莱因,(美)洛普

页数：479

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

## <<Python自然语言处理>>

### 前言

This is a book about Natural Language Processing. By "natural language" we mean a language that is used for everyday communication by humans; languages such as English, Hindi, or Portuguese. In contrast to artificial languages such as programming languages and mathematical notations, natural languages have evolved as they pass from generation to generation, and are hard to pin down with explicit rules. We will take Natural Language Processing—or NLP for short—in a wide sense to cover any kind of computer manipulation of natural language. At one extreme, it could be as simple as counting word frequencies to compare different writing styles. At the other extreme, NLP involves "understanding" complete human utterances, at least to the extent of being able to give useful responses to them. Technologies based on NLP are becoming increasingly widespread. For example, phones and handheld computers support predictive text and handwriting recognition; web search engines give access to information locked up in unstructured text; machine translation allows us to retrieve texts written in Chinese and read them in Spanish. By providing more natural human-machine interfaces, and more sophisticated access to stored information, language processing has come to play a central role in the multi-lingual information society. This book provides a highly accessible introduction to the field of NLP. It can be used for individual study or as the textbook for a course on natural language processing or computational linguistics, or as a supplement to courses in artificial intelligence, text mining, or corpus linguistics. The book is intensely practical, containing hundreds of fully worked examples and graded exercises.

## <<Python自然语言处理>>

### 内容概要

《Python自然语言处理（影印版）》提供了非常易学的自然语言处理入门介绍，该领域涵盖从文本和电子邮件预测过滤，到自动总结和翻译等多种语言处理技术。

在《Python自然语言处理（影印版）》中，你将学会编写Python程序处理大量非结构化文本。你还将通过使用综合语言数据结构访问含有丰富注释的数据集，理解用于分析书面通信内容和结构的主要算法。

《Python自然语言处理》准备了充足的示例和练习，可以帮助你：从非结构化文本中抽取信息，甚至猜测主题或识别“命名实体”；分析文本语言结构，包括解析和语义分析；访问流行的语言学数据库，包括WordNet和树库（treebank）；从多种语言学 and 人工智能领域中提取的整合技巧。

《Python自然语言处理（影印版）》将帮助你学习运用Python编程语言和自然语言工具包（NLTK）获得实用的自然语言处理技能。

如果对于开发Web应用、分析多语言新闻源或记录濒危语言感兴趣——即便只是想从程序员视角观察人类语言如何运作，你将发现《Python自然语言处理》是一本令人着迷且极为有用的好书。

## <<Python自然语言处理>>

### 作者简介

伯德 (Steven Bird) 是墨尔本大学计算机科学和软件工程系副教授, 以及宾夕法尼亚大学语言数据联合会高级研究助理。

克莱因 (Ewan Klein) 是爱丁堡大学信息学院语言技术教授。

洛普 (Edward Loper) 最近从宾夕法尼亚大学获得机器学习自然语言处理博士学位, 目前是波士顿BBN Technologies公司的研究员。

## &lt;&lt;Python自然语言处理&gt;&gt;

## 书籍目录

Preface  
1. Language Processing and Python  
1.1 Computing with Language : Texts and Words  
1.2 A Closer Look at Python : Texts as Lists of Words  
1.3 Computing with Language : Simple Statistics  
1.4 Back to Python : Making Decisions and Taking Control  
1.5 Automatic Natural Language Understanding  
1.6 Summary  
1.7 Further Reading  
1.8 Exercises  
2. Accessing Text Corpora and Lexical Resources  
2.1 Accessing Text Corpora  
2.2 Conditional Frequency Distributions  
2.3 More Python : Reusing Code  
2.4 Lexical Resources  
2.5 WordNet  
2.6 Summary  
2.7 Further Reading  
2.8 Exercises  
3. Processing Raw Text  
3.1 Accessing Text from the Web and from Disk  
3.2 Strings : Text Processing at the Lowest Level  
3.3 Text Processing with Unicode  
3.4 Regular Expressions for Detecting Word Patterns  
3.5 Useful Applications of Regular Expressions  
3.6 Normalizing Text  
3.7 Regular Expressions for Tokenizing Text  
3.8 Segmentation  
3.9 Formatting : From Lists to Strings  
3.10 Summary  
3.11 Further Reading  
3.12 Exercises  
4. Writing Structured Programs  
4.1 Back to the Basics  
4.2 Sequences  
4.3 Questions of Style  
4.4 Functions : The Foundation of Structured Programming  
4.5 Doing More with Functions  
4.6 Program Development  
4.7 Algorithm Design  
4.8 A Sample of Python Libraries  
4.9 Summary  
4.10 Further Reading  
4.11 Exercises  
5. Categorizing and Tagging Words  
5.1 Using a Tagger  
5.2 Tagged Corpora  
5.3 Mapping Words to Properties Using Python Dictionaries  
5.4 Automatic Tagging  
5.5 N-Gram Tagging  
5.6 Transformation-Based Tagging  
5.7 How to Determine the Category of a Word  
5.8 Summary  
5.9 Further Reading  
5.10 Exercises  
6. Learning to Classify Text  
6.1 Supervised Classification  
6.2 Further Examples of Supervised Classification  
6.3 Evaluation  
6.4 Decision Trees  
6.5 Naive Bayes Classifiers  
6.6 Maximum Entropy Classifiers  
6.7 Modeling Linguistic Patterns  
6.8 Summary  
6.9 Further Reading  
6.10 Exercises  
7. Extracting Information from Text  
7.1 Information Extraction  
7.2 Chunking  
7.3 Developing and Evaluating Chunkers  
7.4 Recursion in Linguistic Structure  
7.5 Named Entity Recognition  
7.6 Relation Extraction  
7.7 Summary  
7.8 Further Reading  
7.9 Exercises  
8. Analyzing Sentence Structure  
8.1 Some Grammatical Dilemmas  
8.2 Whats the Use of Syntax?  
8.3 Context-Free Grammar  
8.4 Parsing with Context-Free Grammar  
8.5 Dependencies and Dependency Grammar  
8.6 Grammar Development  
8.7 Summary  
8.8 Further Reading  
8.9 Exercises  
9. Building Feature-Based Grammars  
9.1 Grammatical Features  
9.2 Processing Feature Structures  
9.3 Extending a Feature-Based Grammar  
9.4 Summary  
9.5 Further Reading  
9.6 Exercises  
10. Analyzing the Meaning of Sentences  
10.1 Natural Language Understanding  
10.2 Propositional Logic  
10.3 First-Order Logic  
10.4 The Semantics of English Sentences  
10.5 Discourse Semantics  
10.6 Summary  
10.7 Further Reading  
10.8 Exercises  
11. Managing Linguistic Data  
11.1 Corpus Structure : A Case Study  
11.2 The Life Cycle of a Corpus  
11.3 Acquiring Data  
11.4 Working with XML  
11.5 Working with Toolbox Data  
11.6 Describing Language Resources Using OLAC Metadata  
11.7 Summary  
11.8 Further Reading  
11.9 Exercises  
Afterword : The Language Challenge  
Bibliography  
NLTK  
Index  
General Index

## 章节摘录

Back in elementary school you learned the difference between nouns, verbs, adjectives, and adverbs. These "word classes" are not just the idle invention of grammarians, but are useful categories for many language processing tasks. As we will see, they arise from simple analysis of the distribution of words in text. The goal of this chapter is to answer the following questions :

1. What are lexical categories, and how are they used in natural language processing ?

2. What is a good Python data structure for storing words and their categories ?

3. How can we automatically tag each word of a text with its word class ?

Along the way, we'll cover some fundamental techniques in NLP, including sequence labeling, n-gram models, backoff, and evaluation. These techniques are useful in many areas, and tagging gives us a simple context in which to present them. We will also see how tagging is the second step in the typical NLP pipeline, following tokenization. The process of classifying words into their parts-of-speech and labeling them accordingly is known as part-of-speech tagging, POS tagging, or simply tagging. Parts-of-speech are also known as word classes or lexical categories. The collection of tags used for a particular task is known as a tagset. Our emphasis in this chapter is on exploiting tags, and tagging text automatically.

## <<Python自然语言处理>>

### 媒体关注与评论

“ 很少有这样一本方法清晰、代码整洁的书来讨论如此高难度的计算机问题……这是学习自然语言处理的入门佳作。

” ——Ken Getz, 资深咨询顾问, MCW Technologies公司

<<Python自然语言处理>>

版权说明

本站所提供下载的PDF图书仅提供预览和简介, 请支持正版图书。

更多资源请访问:<http://www.tushu007.com>