

# 《统计自然语言处理（第2版）》

## 图书基本信息

书名：《统计自然语言处理（第2版）》

13位ISBN编号：9787302319115

10位ISBN编号：7302319111

出版时间：2013-8-1

出版社：清华大学出版社

作者：宗成庆

页数：570

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：[www.tushu111.com](http://www.tushu111.com)

# 《统计自然语言处理（第2版）》

## 内容概要

《中文信息处理丛书：统计自然语言处理（第2版）》全面介绍了统计自然语言处理的基本概念、理论方法和最新研究进展，内容包括形式语言与自动机及其在自然语言处理中的应用、语言模型、隐马尔可夫模型、语料库技术、汉语自动分词与词性标注、句法分析、词义消歧、篇章分析、统计机器翻译、语音翻译、文本分类、信息检索与问答系统、自动文摘和信息抽取、口语信息处理与人机对话系统等，既有对基础知识和理论模型的介绍，也有对相关问题的研究背景、实现方法和技术现状的详细阐述。

《中文信息处理丛书：统计自然语言处理（第2版）》可作为高等院校计算机、信息技术等相关专业的高年级本科生或研究生的教材或参考书，也可供从事自然语言处理、数据挖掘和人工智能等研究的相关人员参考。

## 作者简介

宗成庆，中国科学院自动化所研究员、博士生导师。1998年3月毕业于中国科学院计算技术研究所，获博士学位。1998年5月至2000年4月在中国科学院自动化研究所从事博士后研究，博士后出站以后留在自动化所工作至今。2006年至2013年担任模式识别国家重点实验室副主任。曾于1999年和2001年两次在日本国际电气通信基础技术研究所（ATR）做客座研究员。2004年10月至2005年2月在法国格勒诺布尔（Grenoble）信息与应用数学研究院（IMAG）做短期高访。

主要从事自然语言处理、机器翻译和文本分类等相关技术的研究和教学工作，作为项目负责人主持国家自然科学基金项目、国家“863”项目、国家支撑计划项目和中国科学院国际合作项目等10余项，发表论文100余篇，其中在Computational Linguistics、IEEE Intelligent Systems、IEEE TASLP、Information Sciences和ACL、COLING、EMNLP等权威国际期刊和顶级学术会议上发表论文40余篇，出版专著和译著各一部，获国家发明专利十余项。指导开发的多语言机器翻译系统已经在多个国家特定领域获得实际应用，并多次在国际评测中取得优异成绩。

目前任国际计算语言学委员会（International Committee on Computational Linguistics，ICCL）委员、国际计算语言学学会（ACL）汉语特别兴趣组（SIGHAN）候任主席（Chair-Elec），中国中文信息学会常务理事、中国人工智能学会理事、中国计算机学会中文信息技术专委会副主任，曾于2008年至2012年期间担任亚洲自然语言处理联合会（AFNLP）执行理事。目前还担任国际学术期刊ACM TALIP副主编、IJCPOL副主编、IEEE Intelligent systems编委、Machine Translation编委、JCST编委和《自动化学报》副主编等多种学术职务，曾多次在ACL、COLING和EMNLP等顶级学术会议上担任重要职务。2008年获中国科学院集中教学突出贡献奖，2010年获中国科学院“朱李月华优秀教师”奖。享受政府特殊津贴。

## 书籍目录

### 第1章 绪论

#### 1.1 基本概念

##### 1.1.1 语言学与语音学

##### 1.1.2 自然语言处理

##### 1.1.3 关于“理解”的标准

#### 1.2 自然语言处理研究的内容和面临的困难

##### 1.2.1 自然语言处理研究的内容

##### 1.2.2 自然语言处理涉及的几个层次

##### 1.2.3 自然语言处理面临的困难

#### 1.3 自然语言处理的基本方法及其发展

##### 1.3.1 自然语言处理的基本方法

##### 1.3.2 自然语言处理的发展

#### 1.4 自然语言处理的研究现状

#### 1.5 本书的内容安排

### 第2章 预备知识

#### 2.1 概率论基本概念

##### 2.1.1 概率

##### 2.1.2 最大似然估计

##### 2.1.3 条件概率

##### 2.1.4 贝叶斯法则

##### 2.1.5 随机变量

##### 2.1.6 二项式分布

##### 2.1.7 联合概率分布和条件概率分布

##### 2.1.8 贝叶斯决策理论

##### 2.1.9 期望和方差

#### 2.2 信息论基本概念

##### 2.2.1 熵

##### 2.2.2 联合熵和条件熵

##### 2.2.3 互信息

##### 2.2.4 相对熵

##### 2.2.5 交叉熵

##### 2.2.6 困惑度

##### 2.2.7 噪声信道模型

#### 2.3 支持向量机

##### 2.3.1 线性分类

##### 2.3.2 线性不可分

##### 2.3.3 构造核函数

### 第3章 形式语言与自动机

#### 3.1 基本概念

##### 3.1.1 图

##### 3.1.2 树

##### 3.1.3 字符串

#### 3.2 形式语言

##### 3.2.1 概述

##### 3.2.2 形式语法的定义

##### 3.2.3 形式语法的类型

##### 3.2.4 CFG识别句子的派生树表示

## 3.3 自动机理论

### 3.3.1 有限自动机

### 3.3.2 正则文法与自动机的关系

### 3.3.3 上下文无关文法与下推自动机

### 3.3.4 图灵机

### 3.3.5 线性界限自动机

## 3.4 自动机在自然语言处理中的应用

### 3.4.1 单词拼写检查

### 3.4.2 单词形态分析

### 3.4.3 词性消歧

## 第4章 语料库与语言知识库

### 4.1 语料库技术

#### 4.1.1 概述

#### 4.1.2 语料库语言学的发展

#### 4.1.3 语料库的类型

#### 4.1.4 汉语语料库建设中的问题

#### 4.1.5 典型语料库介绍

.....

## 第5章 语言模型

## 第6章 概率图模型

## 第7章 自动分词、命名实体识别与词性标注

## 第8章 句法分析

## 第9章 语义分析

## 第10章 篇章分析

## 第11章 统计机器翻译

## 第12章 语音翻译

## 第13章 文本分类与情感分类

## 第14章 信息检索与问答系统

## 第15章 自动文摘与信息抽取

## 第16章 口语信息处理与人机对话系统

## 参考文献

## 自然语言处理及其相关领域的国际会议

## 名词术语索引

## 精彩短评

- 1、学到很多术语的中文
- 2、比较一般，大量资料罗列导致脉络模糊不清，可以适当精简
- 3、大而全的概要，适合入门
- 4、作为综述来讲已经很不错了
- 5、本科时读过几页就丢了，这学期计算语言学课又重新读的
- 6、对NLP无爱啊
- 7、蛮好，适合入门
- 8、很全的一本NLP中文教材
- 9、浏览了一下，一般。就是作者把NLP的东西嚼完了、整理好了再吐给你。没有深入的见解，没有直觉上的启发。
- 10、大篇幅讲历史，看起来太费劲，不够通俗易懂，本来期望蛮高的，太让我失望了！
- 11、读了一半吧，现在改了方向了，所以暂时搁置。
- 12、读研后我才明白原来导师以为我是万能的#点新技能点中#
- 13、自然语言综述

# 《统计自然语言处理（第2版）》

## 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:[www.tushu111.com](http://www.tushu111.com)