

<<统计自然语言处理>>

图书基本信息

书名：<<统计自然语言处理>>

13位ISBN编号：9787302165989

10位ISBN编号：730216598X

出版时间：2008-5

出版时间：清华大学出版社

作者：宗成庆

页数：475

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>

## <<统计自然语言处理>>

### 内容概要

本书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、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。

本书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。

## <<统计自然语言处理>>

### 作者简介

宗成庆，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模式识别国家重点实验室从事博士后研究，博士后出站以后留在模式识别国家重点实验室工作至今。  
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）

## &lt;&lt;统计自然语言处理&gt;&gt;

## 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
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 语料库与词汇知识库	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 典型语料库介绍
	4.1.5 汉语语料库建设中的问题	4.2 词汇知识库	.....	第5章 语言模型
	第6章 隐马尔可夫模型	第7章 汉语自动分词与词性标注	第8章 句法分析	第9章 语义消歧
	第10章 统计机器翻译	第11章 语音翻译	第12章 文本分类	第13章 信息检索与问答系统
	第14章 自动文摘与信息抽取	第15章 口语信息处理与人机对话系统	附录 项目作业名词术语索引参考文献	

## &lt;&lt;统计自然语言处理&gt;&gt;

## 章节摘录

第1章 绪论 1.1 基本概念 1.1.1 语言学与语音学 我们知道,语言作为人类特有的用来表达情感,交流思想的工具,是一种特殊的社会现象,由语音、词汇和语法构成。语音和文字是构成语言的两个基本属性,语音是语言的物质外壳,文字则是记录语言的书写符号系统〔黄伯荣等,1991〕。

根据《现代语言学词典》〔克里斯特尔,2002〕的定义,语言学(linguistics)是指对语言的科学研究。

作为一门纯理论的学科,语言学在近期获得了快速发展,尤其从20世纪60年代起,已经成为一门知晓度很高的广泛教授的学科。

根据语言学家的注意中心和兴趣范围,语言学可以区分为一些不同的分支,例如,历时语言(diachronic linguistics)或称历史语言学(historical linguistics)、共时语言学(synchronic linguistics)、一般语言学(descriptive linguistics)、理论语言学(theoretical linguistics)、描述语言学(descriptive linguistics)、对比语言学(contrastive linguistics)或类型语言学(typological linguistics)、结构语言学(structural linguistics)等。

语音学(phonetics)是研究人类发音特点,特别是语音发音特点,并提出各种语音描述、分类和转写方法的科学。

语音学一般有三个分支:发音语音学(articulatory),研究发音器官是如何产生语音的;声学语音学(acoustic phonetics),研究口耳之间传递语音的物理属性;听觉语音学(auditory phonetics),研究人通过耳、听觉神经和大脑对语音的知觉反应。

仪器语音学(instrumendal phonetics)则是利用各种物理设备,如测量气流或分析声波的仪器等,来研究上述三个问题的任一方面〔克里斯特尔,2002〕。

## <<统计自然语言处理>>

### 编辑推荐

《中文信息处理丛书·统计自然语言处理》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。

<<统计自然语言处理>>

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>