

<<中文文本信息处理的原理与应用>>

图书基本信息

书名：<<中文文本信息处理的原理与应用>>

13位ISBN编号：9787302154983

10位ISBN编号：7302154988

出版时间：2007-9

出版时间：清华大学出版社

作者：苗夺谦，卫志华 编著

页数：321

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>

<<中文文本信息处理的原理与应用>>

内容概要

本书是一本全面系统地介绍中文文本信息处理的教材，内容丰富，由浅入深地讲述了中文文本信息处理的原理与应用。

本书不仅介绍了基于规则的自然语言分析方法，也介绍了基于统计学的方法。

全书共分为四大部分，分别是词法分析、语法处理、语义分析和应用与技术。

其中前三部分是自然语言处理的基本理论，第一部分针对中文处理中特有的分词问题，介绍了自动分词算法、分词中歧义的消除和未登录词的识别算法，另外还介绍了语料库的相关知识。

第二部分和第三部分都是从语法（语义）的表示入手，将自然语言形式化，再给出语法（语义）分析的算法，并针对该过程中的歧义问题给出了一些成熟的解决方案。

最后一部分讲述自然语言理解在信息检索、信息抽取、自动文摘和文本分类等领域的应用。

本书思路清晰，在每部分及每章的开始都介绍了该部分知识与其他部分之间的关系，以及该部分的知识点之间的关系，以帮助读者从整体上把握中文文本信息处理的思路，并能根据不同的需求或不同的问题选择适当的算法。

本书涉及内容广泛，能满足不同水平读者群的需求，可以作为计算机、信息类高年级本科生的教材，也可作为自然语言处理方向研究生的教材，也非常适合作为自然语言处理应用领域的研究人员和技术人员的参考资料。

<<中文文本信息处理的原理与应用>>

书籍目录

第1章 概论	1.1 自然语言处理与中文信息处理	1.2 自然语言处理的新趋势	1.3 本书内容组织
第一部分 词法分析	第2章 自动分词	2.1 关于自动分词	2.2 分词词典
方法	第3章 分词歧义消解	3.1 关于分词歧义	3.2 基于规则的分词消歧
		3.3 基于统计方法的分词消歧	
第4章 未登录词获取	4.1 关于未登录词	4.2 基于统计学的未登录词获取方法	
4.3 中文姓名的自动辨识	4.4 中文统计词汇获取	4.5 无词典分词方法	
第5章 语料库的构建	5.1 关于语料库	5.2 汉语语料库的基本加工规范	5.3 建设语料库的其他问题
第一部分习题	第二部分 语法处理	第6章 自动标注	6.1 关于自动标注
		6.2 马尔可夫模型和隐马尔可夫模型	6.3 马尔可夫模型标注器
		6.4 隐马尔可夫模型标注器	第7章 语法表示
		7.1 关于语法表示	7.2 形式语法描述
		7.3 短语结构语法	7.4 转移网络
		7.5 短语结构与句法树	第8章 语法分析
		8.1 关于语法分析	8.2 基于符号串的句法分析
		8.3 自底向上的图句法分析	8.4 自顶向下的图句法分析
		8.5 基于转移网络的句法分析	8.6 移进归约句法分析器
		8.7 概率上下文无关文法分析	第二部分习题
第9章 语义表示	9.1 关于语义表示	9.2 语义的逻辑表示方法	9.3 论旨角色
语义网络表示法	9.4 语义网络表示法	9.5 框架表示法	9.6 量词的处理
第10章 语义分析	10.1 关于语义分析	10.2 组合理论与语义解释	10.3 基于语义特征的解释方法
		10.4 基于语法关系的语义分析	10.5 语义语法
		10.6 模板匹配	10.7 语义驱动的分析技术
第11章 语义消歧	11.1 关于语义歧义	11.2 选择限制法消歧	11.3 语义网络
		11.4 统计词义消歧	11.5 统计语义优选
第三部分习题	第四部分 应用与技术	第12章 文本分类	12.1 关于文本分类
		12.2 文本分类方法	12.3 文本分类的评测
		第13章 信息检索	13.1 关于信息检索
		13.2 基于统计学的信息检索模型	13.3 基于语义的信息检索
		13.4 典型信息检索系统	13.5 信息检索技术前沿
第14章 信息抽取	14.1 关于信息抽取	14.2 半结构化文本的信息抽取技术	14.3 典型信息抽取系统
		14.4 Web信息抽取	第15章 自动文摘
		15.1 关于自动文摘	15.2 自动文摘的方法
		15.3 自动文摘系统的评测	15.4 自动文摘系统
第四部分习题	参考文献		

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>