

<<中文信息处理技术教程>>

图书基本信息

书名：<<中文信息处理技术教程>>

13位ISBN编号：9787302117612

10位ISBN编号：7302117616

出版时间：2005-9

出版时间：第1版 (2005年9月1日)

作者：朱巧明

页数：291

字数：436000

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>

<<中文信息处理技术教程>>

内容概要

计算机中文信息处理技术是一门综合性的交叉学科，它包含了计算机科学、数学、认知心理学和语言学等多门学科。

全书分三个部分，比较系统地介绍了计算机中文信息处理技术的研究对象、研究方法和技术。

第1章~第4章主要介绍中文信息处理技术中的一些基础理论和基础知识；第5章~第8章主要介绍中文信息处理技术中的基本I/O技术以及相关的字形技术；第9章~第12章主要介绍了中文信息处理技术中的自然语言处理技术。

本书可以作为高等学校计算机科学与技术专业本科教材或参考书，也可以作为计算机中文信息处理技术研究领域的研究生的教学参考书和计算机技术研发人员的参考书。

<<中文信息处理技术教程>>

书籍目录

第1章 概论 1.1 信息处理 1.2 中文信息处理 1.3 计算机中文信息处理主要研究对象 1.4 现代汉语的特点 1.5 中文信息处理的国际化和本地化 思考题 第2章 汉字信息在计算机内部的表示 2.1 英文字符在计算机内部的表示 2.2 中文信息在计算机中的表示 2.3 汉字代码体系 2.4 汉字的排序 2.5 汉字编码字集 思考题 第3章 汉字字符编码集的应用 3.1 汉字的简繁转换 3.1.1 汉字简化字标准介绍 3.1.2 简—繁转换中遇到的问题 3.1.3 汉字简繁—繁简转换的实现 3.2 Internet上常用的字符编码集 3.2.1 UUENCODE 编码 3.2.2 QUOTED-PRINTABLE编码 3.2.3 BASE64编码 3.2.4 HZ编码 3.2.5 UTF编码 思考题 附录 第4章 汉字编码技术 4.1 汉字编码的发展 4.2 汉字编码中的几个概念 4.2.1 字汇和词汇 4.2.2 码元 4.2.3 码长 4.2.4 单码和重码 4.2.5 编码空间和编码效率 4.3 汉字编码理论 4.3.1 汉字的熵 4.3.2 汉字键盘编码的依据 4.3.3 汉字编码分类 4.3.4 键盘编码和键盘 4.4 数码键盘方案 4.4.1 纵横码 4.4.2 五笔数码 4.4.3 统一码 4.5 汉字编码的标准和规范 4.5.1 汉字编码国家标准和规范 4.5.2 与汉字编码有关的标准与规范说明 4.5.3 汉字键盘编码和输入系统的性能指标 思考题 第5章 中文平台与中文操作系统 5.1 中文平台与中文操作系统历史 5.2 中文操作系统设计方法 5.3 汉字处理模块与汉字代码体系 5.4 常用中文操作系统简介 5.4.1 CC-DOS 5.4.2 中文Windows 5.4.3 中文Linux操作系统 5.4.4 中文嵌入式操作系统 思考题 第6章 汉字输入技术 6.1 汉字输入技术概述 6.1.1 汉字输入发展史 6.1.2 汉字输入分类 6.1.3 汉字输入技术的发展和应 6.2 汉字键盘输入技术 6.2.1 键盘 6.2.2 汉字键盘输入原理 6.2.3 键盘汉字输入的总 6.2.4 输入码对照表的设计 6.3 汉字键盘输入系统 6.3.1 汉字输入系统的分类 6.3.2 汉字输入系统的功能 6.3.3 汉字键盘输入系统的工作流程 6.4 Windows汉字输入技术 6.4.1 概述 6.4.2 Windows中的汉字输入原理 6.4.3 输入法生成器 6.4.4 在SDK和DDK环境下开发输入系统 6.5 Linux下的汉字输入系统 6.5.1 概述 6.5.2 XIM协议 6.5.3 输入法服务器和输入法 6.5.4 输入法设计 思考题 第7章 汉字字形和字形库管理技术 7.1 汉字字形概述 7.1.1 汉字字形及其特点 7.1.2 字形、字型和字形库 7.1.3 字形描述技术及其种类 7.2 字形描述技术 7.2.1 点阵字形描述技术 7.2.2 轮廓矢量字形描述技术 7.2.3 曲线轮廓字形描述技术 7.2.4 其他字形描述技术 7.3 字形的压缩和还原 7.3.1 字形点阵的压缩和还原 7.3.2 轮廓矢量字形的压缩方法 7.3.3 其他字形压缩技术 7.4 字形的放大和缩小 7.4.1 汉字的字号 7.4.2 字形放大和缩小的基本原理 7.4.3 防止字形放大和缩小失真措施 7.4.4 字形放大缩小的应用 7.5 TTF字形技术 7.5.1 什么是TrueType 7.5.2 TrueType的基本原理 7.5.3 TrueType字体文件结构 7.5.4 TrueType的特点和优势 7.5.5 TrueType的应用 7.5.6 OpenType的出现 7.6 字库的设计和管理技术 7.6.1 汉字字库的基本要求 7.6.2 汉字库性能的评测 7.6.3 汉字库结构的设计 7.6.4 点阵字库的制作 7.7 汉字字库的应用 7.7.1 造字程序的使用 7.7.2 安装新字体 思考题 第8章 汉字输出技术 8.1 汉字的输出技术综述 8.2 汉字输出原理和基本结构 8.2.1 显示设备 8.2.2 汉字显示的原理 8.2.3 内码缓冲区和字形缓冲区 8.2.4 汉字显示输出过程 8.2.5 汉字终端 8.3 汉字显示技术实例 8.3.1 DOS下的汉字显示技术 8.3.2 Windows的汉字显示技术 8.4 打印输出原理和基本结构 8.4.1 打印设备 8.4.2 汉字打印概述 8.4.3 汉字打印原理 8.4.4 汉字打印过程 8.4.5 汉字打印机 思考题 第9章 汉语分词 9.1 汉语分词的概念 9.2 分词词典 9.2.1 分词词典简介 9.2.2 基于词属性的分词词典 9.2.3 基于二次索引的分词词典 9.3 常用的汉语分词算法 9.3.1 正向最大匹配算法 9.3.2 逆向最大匹配算法 9.3.3 邻近匹配算法 9.3.4 最短路径匹配算法 9.3.5 基于统计的最短路径分词算法 9.4 无词典分词方法 9.4.1 分词模型 9.4.2 无词典分词算法 9.5 交集型歧义的切分 9.5.1 基本概念 9.5.2 利用互信息和t-测试差处理交集型歧义切分 思考题 第10章 中文信息检索 10.1 信息检索概述 10.1.1 信息检索的定义 10.1.2 结构、半结构和非结构化文档 10.1.3 信息检索的任务 10.1.4 信息检索的评测 10.1.5 中文信息检索的特点 10.1.6 信息检索的模型 10.1.7 信息检索系统的结构 10.2 基于统计的信息检索模型 10.2.1 传统的布尔模型 10.2.2 发展的布尔模型 10.2.3 向量空间模型 10.2.4 概率模型 10.3 基于语义的信息检索 10.3.1 自然语言处理 10.3.2 潜在语义索引 10.3.3 神经网络 10.4 信息检索系统介绍 10.5 Web信息检索 10.5.1 Web信息检索系统 10.5.2 Web信息检索的特点 10.5.3 搜索引擎 10.5.4 搜索引擎的网页分级算法 10.5.5 常用Web搜索引擎 10.5.6 搜索引擎技术的发展趋势 思考题 第11章 中文信息抽取 11.1 信息抽取概述 11.1.1 信息抽取的含义 11.1.2 信息抽取实例 11.1.3 信息抽取和信息检索 11.1.4 中文信息抽取的特

<<中文信息处理技术教程>>

殊性 11.1.5 信息抽取的历史和现状 11.1.6 信息抽取系统的评测 11.1.7 信息抽取任务 11.2 信息抽取系统的结构 11.2.1 信息抽取系统的构建方法 11.2.2 通用信息抽取结构 11.2.3 Bare Bones结构 11.3 信息抽取中的自然语言处理技术 11.3.1 命名实体识别 11.3.2 句法分析 11.3.3 文章分析和推理 11.3.4 知识理解 11.4 信息抽取技术 11.4.1 基于规则的信息抽取技术 11.4.2 隐马尔可夫模型 11.5 Web信息抽取 11.5.1 基于自然语言处理方式的信息抽取 11.5.2 包装器方式的信息抽取 11.5.3 基于本体方式的信息抽取 11.5.4 基于HTML结构的信息抽取 11.5.5 基于Web查询的信息抽取 思考题 第12章 中文文本分类技术 12.1 文本分类的概念 12.1.1 什么是文本分类 12.1.2 文本分类的两种类型 12.1.3 文本分类的两种方式 12.1.4 中文文本分类 12.1.5 文本分类的国内外研究情况 12.2 文本特征的选择 12.2.1 预处理 12.2.2 文本特征的选择 12.3 文本分类方法 12.3.1 Rocchio方法——相似度计算方法 12.3.2 Na[?]ve Bayes——贝叶斯方法 12.3.3 KNN方法——K近邻方法 12.3.4 SVM——支持向量机 12.3.5 Decision Tree——决策树方法 12.4 选择阈值的策略 12.5 评估方法 思考题 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>