

图书基本信息

书名：<<情报检索语言与智能信息处理丛书 (全8册) >>

13位ISBN编号：9787564119133

10位ISBN编号：7564119136

出版时间：2009-12

出版时间：东南大学出版社

作者：章成志，白振田 著

页数：198

字数：173000

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>

前言

2001年7-8月《中图法》编委会在京对《中国分类主题词表》第2版初稿进行综合审定，南京农业大学、中山大学、上海空军政治学院等信息管理系的老师带领部分研究生参加了审定工作。

薛春香就是这批学生中的佼佼者之一。

她敏捷严谨的思维、沉稳扎实的作风以及深究勤问的性格给我留下深刻的印象。

此后，她攻读博士、做博士后，始终在信息组织这块沃土上辛勤耕耘，由此也与《中图法》、《中国分类主题词表》结下了不解之缘，成为《中图法》编委会的常客。

八年后，薛春香的《网络环境中知识组织系统构建与应用研究》问世，这是她在博士论文的基础上从更广阔的情报语言领域进行深入研究后完成的，论证的结构更加完善、内容也更加丰富。

这本书对网络环境中的知识组织系统及研究现状、知识组织系统构建与描述标准等作了全面的概述，深入对网络环境中知识组织系统的设计、传统知识组织系统自动构建以及知识组织系统的互操作与整合进行了研究，并对网络环境中知识组织系统的应用做了分析，是目前国内一本较全面地研究知识组织系统构建与应用的著作。

内容概要

本书总结了作者近年来在文本自动标引和自动分类上所做的研究与实践。

全书由四部分构成：第一部分为基础部分，介绍研究背景和研究意义，并对相关研究进展进行综述；第二部分介绍作者在自动标引方面的研究工作；第三部分介绍基于《中图法》分类知识库的文本自动分类系统；第四部分介绍基于统计与规则相结合的文本自动分类系统。

本书是国内第一本比较系统地介绍文本自动标引和自动分类研究的专著，对从事信息检索、文本挖掘、知识组织、数字图书馆等研究和应用开发的科技人员有较高参考价值，可作为图书馆学、情报学、计算机科学与技术、信息管理和信息系统等专业的研究生、高年级本科生的教学参考书和技术资料。

作者简介

章成志，1997年生，1999年获安徽工程科技学院学士学位，2002年获南京农业大学硕士学位，2007年获南京大学博士学位。

现为中国科学技术信息研究所在站博士后，南京理工大学信息管理系讲师、硕士生导师。

研究领域包括信息组织、信息检索、文本挖掘等。

曾获国家自然科学基金

书籍目录

第一部分 第1章 引言	1.1 研究背景	1.2 自动标引与自动分类的作用	1.3 本书的内容与章节安排
第2章 文本自动标引与分类研究进展	2.1 自动标引研究综述	2.2 文本分类研究综述	2.3 本章小结
第二部分 第3章 文本分词技术及抽词词典构造	3.1 文本分词技术概述	3.2 分词模式设计及其原理	3.3 原始抽词词典的构造
3.4 词典约简算法实验	第4章 基于多特征选择及权重计算	4.1 特征选择方法概述	4.2 算法设计原理
4.3 结果分析	第5章 自动标引中标引源权重方案确定	5.1 标引源权重研究综述	5.2 标引源权重方案的确定
5.3 本章小结	第三部分 第6章 分类知识库的制作	6.1 概述	6.2 关键词(串)一分类号关联研究综述
6.3 关键词(串)一分类号关联方法	6.4 分类知识库的制作	6.5 分类知识库的性能测评	6.6 篇名知识库的制作
6.7 本章小结	第7章 基于语义体系的词语相似度计算	7.1 概述	7.2 词语相似度研究综述
7.3 基于语义体系的词语相似度算法	7.4 基于语义相似度的同义词挖掘	7.5 本章小结	第8章 基于知识库的文本自动分类
8.1 文本自动系统总体设计	8.2 文本自动分类系统的测评	8.3 《全国报刊索引》自动标引与自动分类系统介绍	8.4 本章小结
第四部分 第9章 统计与决策规则双重分类算法	9.1 分类器概述	9.2 双重分类原理	9.3 分类规则提取
9.4 双重分类过程	9.5 实验结果及分析	第10章 层次分类算法实验	10.1 层次分类原理
10.2 层次分类算法设计	10.3 实验结果及分析	第11章 基于统计与规则相结合的文本分类系统的实现	11.1 系统实验用材料选择及分析
11.2 系统总体框架与模块介绍	11.3 系统测试分析	11.4 本章小结	名称索引 主题索引 后记

章节摘录

插图：(3) 术语表和字典 / 词典术语表是术语及其定义的列表。

术语可来源于特定领域或特定作品，这些术语在某一特定环境中被定义，而且很少有多种不同的含义。

字典 / 词典是以字顺排列的字或词以及它们的定义的列表，包括字或词的各种不同含义。

字典 / 词典的范围要比术语表宽泛和通用。

字典 / 词典还可以提供字或词的起源信息，如拼写或词法，以及跨学科的多重含义；有的字典或词典还可能提供不同含义的同义词、相关词，但它没有明确的等级层次体系，也没有试图通过概念来将字词加以分组，如《现代汉语词典》等。

(4) 规范文档规范文档亦称权威档，是用来控制同一实体（人或事物）不同的名称，或者某个特定领域中专有名词的词汇表，例如国家名称表、个人姓名表、组织结构名称列表。

非规范的词汇被链接指向规范词汇。

这种类型的知识组织系统并不揭示复杂的结构和组织关系，一般只揭示等同关系。

如台湾“中央研究院”组织制作的人名权威档、地名权威档、时代权威档等，在其人名权威档中检索“徐光启”，在“徐光启”这一条目下，以“徐光启”作为规范名，并列有其字、号，如“子先”、“徐玄扈”、“玄扈先生”、“徐文定公”等，通过这些词汇的检索能指向“徐光启”这一条目。

(5) 同义词环同义词环是一组同义、近义甚至相关词的集合，类似于入口词表，但是它不会强行要求环中某一个词作为“正式词”，环中任何一个词的地位都是平等的，都可以是正式词。

同义词环通常与搜索引擎一起使用，以帮助扩充检索式，用户只要使用同义词环中的任何一个词，就可以找到这一类的资源。

图1-5是利用Googl。

后记

本书系《情报检索语言与智能信息处理》丛书之一，是在笔者博士学位论文《农史知识组织系统构建与应用研究》（2006年6月完成）的基础上经过扩充改编而成。

知识组织系统是对各种知识组织的语义工具统称。

从广义而言，其包含的各种具体类型的工具有数十种之多，本书在论述时尽量从整体上论述，以求全面。

但限于笔者以往的研究以分类法和词表建设与应用为主，对本体、主题图等缺乏实践；同时，分类法和受控词表仍是本体、主题图等新型知识组织系统构建的基础，因此在本书实验研究中仍以网络环境中的分类法和受控词表为研究重点。

但是从传统分类法、叙词表向本体转换，基于概念图和主题图来组织和导航信息资源等，将是笔者今后研究中不可回避并将着重关注的课题。

本书能够得以出版，首先感谢我的导师侯汉清教授，没有导师的亲切教诲和悉心指导，要完成本书是不可想象的。

导师一贯严谨的治学作风、孜孜不倦的探索精神给我留下深切的体会，并鞭策我在今后的人生旅程中不断学习与提高。

编辑推荐

《网络环境中知识组织系统构建与应用研究》：情报检索语言与智能信息处理丛书

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>