<<信息检索系统导论>>

图书基本信息

书名：<<信息检索系统导论>>

13位ISBN编号：9787111246077

10位ISBN编号：7111246071

出版时间：2008-12

出版时间：机械工业出版社

作者：刘挺 等编著

页数：257

前言

信息检索和搜索引擎因Internet的普及而日益变成一个热门学科。
各种相关学科的技术都被用于信息检索，而信息检索也被用于各个领域。
热门固然是一门学科兴盛的表现，每个从事研究的人都希望自己的研究领域成为热门。
但热门也可能带来危险，即把信息检索当作一种时髦技术，无论适用与否都将其套用而不究其根本。
对于信息检索而言，这种时髦反而是它进一步发展的障碍。

<<信息检索系统导论>>

内容概要

本书对信息检索及信息检索系统的基本概念、原理、算法进行详尽介绍。
主要内容包括信息检索模型、文本操作技术、文本索引和搜索技术、查询处理与Web检索技术、分布式信息检索、文本分类与聚类、信息过滤等，并给出Web信息检索的实现实例。
本书内容丰富，源于作者多年的教学及科研心得，适合作为高等院校计算机专业本科生及研究生相关课程的教材，也可作为技术人员研究信息检索与搜索引擎的参考读物。

作者简介

刘挺，教授，博士生导师。
哈尔滨工业大学计算机研究所副所长，信息检索研究室主任。
国家863”中文处理”重点项目总体组专家。
中国中文信息学会理事．信息检索专委会副主任．计算语言学专委会委员，《中文信息学报》编委。
中国计算机学会中文信息技术专委会委员，YOCSEF委员。

<<信息检索系统导论>>

书籍目录

主要研究内容 11.2.1 信息过滤的概念和主要特点 11.2.2 信息过滤与信息检索、信息抽取以及分类等研究的区别 11.2.3 信息过滤系统的分类体系 11.3 信息过滤系统的结构及评价 11.3.1 信息过滤系统的组成 11.3.2 信息过滤系统的评价 11.4 基于内容的信息过滤 11.4.1 信息过滤中应用的统计模型 11.4.2 信息过滤中应用的文本分类方法 11.5 协作过滤 11.5.1 基于用户的协作过滤 11.5.2 基于模型的协作过滤 11.5.3 基于项目的协作过滤 本章小结 思考练习 参考文献第12章 问答系统 12.1 引言 12.2 问答系统的发展历程 12.3 问答系统的种类 12.3.1 问答系统分类方法 12.3.2 自然语言的数据库问答系统 12.3.3 对话式问答系统 12.3.4 基于常问问题集的问答系统 12.3.5 基于大规模文档集的问答系统 12.3.6 阅读理解系统 12.3.7 基于知识库的问答系统 12.4 基于常问问题集的问答系统实现 12.4.1 候选问题集的建立 12.4.2 句子相似度计算 12.5 基于大规模文档集的问答系统实现 12.5.1 问答的任务与系统实现流程 12.5.2 问题分析 12.5.3 相关文档检索 12.5.4 句段检索 12.5.5 答案抽取 12.5.6 问答结果的答案评测及其面对的问题和困难 本章小结 思考练习 参考文献

章节摘录

第1章 绪论1.1 信息检索简介1.1.1 信息检索的概念和处理对象什么是信息检索呢?概括地说，信息检索就是从非结构化的信息集合中找出与用户需求相关的信息。
相应的，信息检索系统就是用来实现信息检索功能的计算机软件系统。
这里要强调的是，与数据库系统处理的结构化信息不同，信息检索系统处理的是“非结构化信息”。
什么是“非结构化信息”呢?一篇新闻就是一条非结构化信息，新闻中会出现一些人名、地名、机构名等实体，以及这些实体之间的关系（比如某人是某地区某机关的负责人），还有与这些实体相关的事件（比如某人访问了某地）。
但是这些人、事、物、关系和事件并不像关系数据库的二维表中存放的信息那样，被精确地分割并严格地存放在合适的字段或记录中。
这种在现实世界中自然存在的模糊而带有歧义且没有经过规格化的信息被称为“非结构化的”（unstructured）信息。
现实世界中存在着大量的非结构化信息，除文本外，还有图像、图形、语音、视频等多媒体信息。
本书不讨论多媒体检索，而是专注于文本检索，因此本书中所涉及的检索对象默认为文本。
文本又有各种各样的类型，如网页、邮件、博客、论坛上的帖子、聊天记录、短信等，不同类型的文本有不同的特点，比如论坛上的帖子往往非常口语化，存在大量的别称、省略语等现象，给检索带来很大的挑战。

<<信息检索系统导论>>

编辑推荐

《信息检索系统导论》特点：理论联系实际，介绍了用Lucene实现Web信息检索的实例。
融入了作者的科研实践心得，对相关的前沿技术也有所涉及。
每章后都附有大量的参考文献，并提供思考题供读者进行深入研究。
《信息检索系统导论》为授课教师提供电子课件，请登录华章网站下载。

版权说明