第一图书网, tushu007.com

<<深入搜索引擎>>

图书基本信息

书名：<<深入搜索引擎>>

13位ISBN编号：9787121084911

10位ISBN编号：7121084910

出版时间：2009

出版时间：电子工业出版社

作者：ian H. Witten,Alistair Moffat,Timothy C. Bell

页数：540

字数：574000

译者：梁斌

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

内容概要

本书是斯坦福大学信息检索和挖掘课程的首选教材之一，并已成为全球主要大学信息检索的主要教材。
本书理论和实践并重，深入浅出地给出了海量信息数据处理的整套解决方案，包括压缩、索引和查询的方方面面。
其最大的特色在于不仅仅满足信息检索理论学习的需要，更重要的是给出了实践中可能面对的各种问题及其解决方法。
本书作为斯坦福大学信息检索课程的教材之一，具有一定的阅读难度，主要面向信息检索专业高年级本科 生和研究生、搜索引擎业界的专业技术人员和从事海量数据处理相关专业的技术人员。

作者简介

Ian H.Witten 是新西兰Waikato大学计算系科学系教授，是ACM、新西兰皇家学会会员。
是英国、美国、加拿大和新西兰的专业计算、信息检索和工程协会会员。
他是 《The Reactive Keyboard》和《Text Compression》的作者之一，这两本书分别出版于1992年和1990年。
各大会议和期刊论文都能看到他的论文。
Alistair Moffat是墨尔本大学计算科学系的副教授。
在各大会议和期刊中发表了大量论文，这些论文包括的领域有：关于文本和图像压缩的算法和数据结构，字典和优先级队列的自适应数据结构，以及自适应搜索和排序算法。
Timothy C.Bell是Canterbury大学计算机科学系系主任。
是出版于1990年的《Text Compression》一书的作者。
在各大期刊和会议上发表了多篇论文，这些论文涉及文本和图像压缩，计算机和音乐，计算机教育等。

第一图书网, tushu007.com

书籍目录

第1章 概览 1.1 文档数据库（DOCUMENT DATABASES） 1.2 压缩（COMPRESSION） 1.3 索引（INDEXES） 1.4 文档索引 1.5 MG海量文档管理系统 1.6 进一步阅读 第2章 文本压缩 2.1 模型 2.2 自适应模型 2.3 哈夫曼编码 范式哈夫曼编码 计算哈夫曼编码长度 总结 2.4 算术编码 算术编码是如何工作的 实现算术编码 保存累积计数 2.5 符号模型 部分匹配预测 块排序压缩 动态马尔科夫压缩 基于单字的压缩 2.6 字典模型 自适应字典编码器的LZ77系列 LZ77的Gzip变体 自适应字典编码器的LZ78系列 LZ78的LZW变体 2.7 同步 创造同步点 自同步编码 2.8 性能比较 压缩性能 压缩速度 其他性能方面的考虑 2.9 进一步阅读 第3章 索引 3.1 样本文档集合 3.2 倒排文件索引 3.3 压缩倒排文件 无参模型（Nonparameterized models） 全局贝努里模型 全局观测频率模型（Global observed frequency model） 局部贝努里模型（Local Bernoulli model） 有偏贝努里模型（Skewed Bernoulli model） 局部双曲模型（Local hyperbolic model） 局部观测频率模型（Local observed frequency model） 上下文相关压缩（Context-sensitive compression） 3.4 索引压缩方法的效果 3.5 签名文件和位图 签名文件 位片签名文件（Bitsliced signature files） 签名文件分析 位图 签名文件和位图的压缩 3.6 索引方法的比较 3.7 大小写折叠、词根化和停用词 大小写折叠 词根化 影响索引长度的因素 停用词（stop word） 3.8 进一步阅读 第4章 查询 4.1 访问字典的方法 访问数据结构 前端编码（Front coding） 最小完美哈希函数 完美哈希函数的设计 基于磁盘的字典存储 4.2 部分指定的查询术语 字符串暴力匹配（Brute-force string matching） 用n-gram索引 循环字典（Rotated lexicon） 4.3 布尔查询（BOOLEAN QUERY） 合取查询（conjunctive query） 术语处理顺序 随机访问和快速查找 分块倒排索引 非合取查询（Nonconjunctive query） 4.4 信息检索和排名 坐标匹配（Coordinate matching） 内积相似度 向量空间模型 4.5 检索效果评价 召回率和精确率 召回率-精确率曲线 TREC项目 208 万维网搜索（World Wide Web Searching） 其他有效性评价方法 4.6 余弦法实现 文档内频率 余弦值的计算方法 文档权重所需的内存 累加器内存 快速查询处理 按频率排序的索引 排序 4.7 交互式检索 相关性反馈 概率模型 4.8 分布式检索 4.9 进一步阅读第5章 索引构造 第6章 图像压缩 第7章 文本图像 第8章 混合图文 第9章 系统实现第10章 信息爆炸 附录A MG系统指南附录B 新西兰图书馆

第一图书网, tushu007.com

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