第一图书网, tushu007.com

<<大规模分布式内容检索技术>>

图书基本信息

书名：<<大规模分布式内容检索技术>>

13位ISBN编号：9787030314178

10位ISBN编号：7030314174

出版时间：2011-5

出版时间：科学出版社

作者：陈汉华，金海　著

页数：295

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

大规模分布式内容检索是近年来分布式系统方向的一个热点研究领域。
陈汉华，金海的这本《大规模分布式内容检索技术》全面地阐述了各种体系结构的分布式大规模内容检索系统的关键技术和核心理论，并对各项技术和理论的来龙去脉进行了详细深入的分析。

本书通过丰富的文献资料和研究成果，从研究者的视角对大规模分布式内容检索技术进行了深入剖析，是分布式处理系统领域的学术专著。
《大规模分布式内容检索技术》可供高等院校计算机科学与技术相关专业的高年级本科生、研究生、教师、研究人员及工程技术人员阅读参考，也可作为相关专业的研究生教材。

书籍目录

序
前言
第1章 绪论
1.1 对等网络概述
1.2 基于对等模式的大规模分布式文本内容检索
1.3 大规模分布式文本内容检索研究面临的挑战
1.4 大规模分布式文本内容检索技术分类
1.4.1 基于结构化分布式哈希表的分布式全局倒排索引
1.4.2 基于非结构化对等网络的联邦式搜索网络
1.4.3 混合对等网络搜索引擎
1.5 本书内容
参考文献
第2章 分布式哈希表及单关键字全局索引
2.1 分布式哈希表
2.1.1 Chord：基于二分查找的环状对等结构
2.1.2 CAN：基于多维空间划分的对等结构
2.1.3 Pastry：基于多分查找的前缀匹配对等结构
2.1.4 Tapestry：基于多分查找的对等结构
2.2 现有分布式哈希表算法的比较
2.3 利用分布式哈希表构建单关键字全局索引
2.3.1 eSearch：基于分布式哈希表的水平索引
2.3.2 Minerva：在查询中挖掘关联关键字
2.3.3 局限性
参考文献
第3章 布隆滤波
3.1 哈希编码的时间/空间权衡
3.1.1 一种经典的哈希编码方法
3.1.2 两种存在误判率的哈希编码方法
3.1.3 计算因子
3.1.4 三种哈希编码方法的数学分析
3.1.5 时空性能比较
3.2 布隆滤波的基本理论
3.2.1 布隆滤波概念
3.2.2 位向量长度的下界
3.2.3 布隆滤波与集合运算
3.3 布隆滤波的扩展形式
3.3.1 计数布隆滤波
3.3.2 压缩布隆滤波
3.3.3 动态布隆滤波
3.4 布隆滤波的应用
3.4.1 早期应用
3.4.2 分布式缓存
3.4.3 P2P网络
3.4.4 资源路由
3.4.5 数据包路由

3.4.6 基础设施测量
参考文献
第4章 基于分布式哈希表单关键字索引的搜索
4.1 结构化对等网多关键字检索面临的挑战
4.2 Top-k查询策略
4.2.1 倒排索引
4.2.2 Top-k裁剪算法
4.2.3 性能评估
4.3 PWEB系统
4.3.1 PWEB网络结构
4.3.2 多关键字搜索通信开销优化策略
4.3.3 扩展性算法
4.3.4 分布式交集运算执行顺序优化策略
4.3.5 搜集关键字全局统计信息
4.3.6 模拟仿真方法
4.3.7 性能评估
4.4 小结
参考文献
第5章 多关键字全局索引及搜索
5.1 分布式关键字集索引面临的挑战
5.2 文本检索中的关键字权重方法
5.2.1 关键字权重模型TF×IDF
5.2.2 理解逆文档频率
5.2.3 用逆向总关键字频率替换逆文档频率的尝试
5.2.4 词频在相关权重模型中的探索
5.3 HDK：基于高区分关键字集的索引技术
5.3.1 关键字集倒排索引
5.3.2 高区分关键字集索引
5.3.3 基于高区分关键字集索引的搜索
5.3.4 扩展性分析
5.3.5 性能评估
5.4 TSS：基于关键字集索引的P2P搜索系统
5.4.1 TSS系统结构
5.4.2 分布式关键字集索引
5.4.3 模拟测试方法
5.4.4 性能评估
参考文献
第6章 基于复制的联邦式对等搜索策略
6.1 理论分析
6.1.1 模型建立
6.1.2 均匀复制策略和比例复制策略
6.1.3 平方根复制策略
6.1.4 混合复制策略
6.1.5 分布式复制算法的实现
6.2 基于随机游走的随机复制策略
6.2.1 生日悖论和理论下界
6.2.2 随机游走复制策略和搜索协议

6.2.3 性能评估
6.3 BubbleStorm：基于随机多图的概率穷尽搜索策略
6.3.1 副本数量的确定
6.3.2 网络大小的测量
6.3.3 随机多图与随机采样
6.3.4 洪泛和随机游走的完美结合
6.3.5 系统分析
6.3.6 性能评估
6.4 BloomCast：基于轻量级分布式哈希表的随机采样
6.4.1 BloomCast网络结构
6.4.2 网络结点数量估计
6.4.3 随机结点采样
6.4.4 基于布隆滤波的复制算法
6.4.5 多关键字搜索
6.4.6 性能评估
6.5 PlanetP：基于全局摘要索引的复制策略
6.5.1 全局目录索引复制
6.5.2 结点排序模型
6.5.3 查询处理算法
6.5.4 性能评估
参考文献
第7章 基于内容路由的联邦式搜索策略
7.1 基于语言模型的路由选择
7.1.1 联邦式搜索引擎的两层结构
7.1.2 语言模型
7.1.3 相对熵
7.1.4 搜索算法
7.1.5 性能评估
7.2 基于语义小世界模型的联邦式对等搜索
7.2.1 语义空间和向量
7.2.2 构造语义小世界
7.2.3 降低语义小世界的维度
7.2.4 基于语义小世界的搜索
7.2.5 性能评估
7.3 基于兴趣局部性的路由
7.3.1 兴趣局部性
7.3.2 基于兴趣局部性的拓扑和路由
7.3.3 性能评估
7.4 SemreX系统
7.4.1 SemreX系统模型
7.4.2 语义覆盖网
7.4.3 基于语义覆盖网的查询搜索算法
7.4.4 性能评估
参考文献
第8章 混合式对等搜索策略
8.1 混合对等搜索面临的挑战
8.2 基于预先探测的混合策略

8.2.1 Boon Thau Loo的Gnutella实验
8.2.2 SimplelHbrid混合P2P搜索策略
8.2.3 性能评估
8.3 基于Gossip的混合搜索选择
8.3.1 收集全局统计信息
8.3.2 使用全局信息进行搜索选择
8.3.3 洪泛阈值的调节
8.3.4 性能评估
8.4 难度感知的混合式搜索策略
8.4.1 很多复本≠很多结点
8.4.2 QRank设计
8.4.3 用QRank进行混合查询
8.4.4 自适应混合查询
8.4.5 QRank仿真器设计
8.4.6 性能评估
参考文献
第9章 大规模在线社会网络搜索
9.1 大规模在线社会网络搜索面临的挑战
9.2 在线社会网络系统研究现状
9.3 流行在线社会网络的数据划分与定位
9.4 大规模在线社会网络内容搜索关键技术
9.4.1 流式文本摘要技术
9.4.2 基于摘要索引的排序算法
9.4.3 多跳邻居摘要聚合技术
9.4.4 基于社区局部性降低摘要索引开销
参考文献

章节摘录

第1章 绪论随着网络技术的迅猛发展和网络应用的迅速普及，互联网日益形成一个巨大的分布式信息库。
互联网应用产生的超大规模信息对现有的网络数据管理基础设施提出了新的更为严峻的挑战。
互联网信息库的无限扩张性和与生俱来的分布式特性使非集中式的数据管理和共享机制成为近年来的研究热点［1］。
大规模分布式内容检索研究具有重要的学术价值和应用价值。
对等计算技术是分布式系统和计算机网络结合的产物，它在网络协议的应用层打破了传统的客户机／服务器（C／S）模式，以自主、平等的原则将处于网络边缘的计算、存储、通信、信息等各种资源有效地共享起来，形成协作网络［2，3］。
自诞生以来的短短几年时间里，对等文件共享应用，如Gnutella［4］、KaZaA［5］、BitTor-rent［6］等都取得了极大的成功，并占据了当前互联网一半以上的网络流量。
对等模式因其可扩展性、鲁棒性和动态自适应性等优点，在大规模互联网应用的数据管理和搜索领域日益展现出巨大的潜力。
近年来，基于对等模式的大规模分布式数据管理和内容搜索系统如雨后春笋般涌现出来［7-13］。
本书将围绕大规模分布式内容检索应用展开，全面系统地阐述大规模分布式内容检索系统的关键理论和支撑技术。

编辑推荐

陈汉华，金海的这本《大规模分布式内容检索技术》围绕大规模分布式内容检索技术展开，通过结合对等网络的概念、结构、资源描述和组织、资源定位和路由选择、结果融合及排序方法等，从结构化对等网络、无结构对等网络、和混合式对等网络各自的特点出发，全面系统的阐述了各种环境下进行分布式大规模内容检索的策略。
包括：基于结构化DHT的分布式全局倒排索引和文本检索策略；联邦式对等搜索策略；混合对等搜索策略。
在每一类方法的讨论中作者结合自己的研究实践对已有工作的来龙去脉进行了详细深入地分析。

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