

<<数据索引与数据组织模型及其应用>>

图书基本信息

书名：<<数据索引与数据组织模型及其应用>>

13位ISBN编号：9787309086942

10位ISBN编号：7309086945

出版时间：2012-7

出版时间：复旦大学出版社

作者：胡运发

页数：204

字数：335000

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>

## <<数据索引与数据组织模型及其应用>>

### 内容概要

胡运发编著的《数据索引与数据组织模型及其应用》是面向海量数据处理的属于数据与知识工程范畴的学术专著。

本书从数据组织的观点处理数据索引问题，提出一种互关联后继索引的商空间模型，从数学变换角度论证该模型优越于现有经典索引模型的多种特性：保序性、保假性和压缩性，导出多种优越能力：

(1) log级的快速查询能力；(2) 原文生成能力；(3) 高度的压缩能力。

本书以互关联后继索引模型为中心，与多种领域的核心问题相结合，说明该模型如何对数据压缩、全文数据库、关系数据库、Web数据库、演绎数据库、知识库、文本信息隐藏、数据挖掘等领域产生创新性的影响，从而生成一系列新成果。

《数据索引与数据组织模型及其应用》对从事关系数据库、Web数据库、事务库、演绎数据库、知识库、逻辑语言、搜索引擎、数据云、数据压缩、文本分类、数据挖掘等领域的研究人员、工程技术人员、高等院校的教师与学生有重要的参考价值或借鉴作用。

## 作者简介

胡运发，1964年复旦大学数学系毕业，后任该校计算机系教授，博士生导师。研究方向是数据、知识工程与知识库、创建新型数据索引-数据组织模型。曾参加我国银河亿次计算机的研制工作、逻辑语言Prolog开发、智能计算机的研制，承担“八六三”项目六次，国家自然科学基金项目四次，参加国家自然科学基金重点项目两次。期间获得电子工业科技进步二等奖（1986）、国防科工委科技进步一等奖（1994）及二等奖两次、中国人民解放军三等功一次。上海科技进步奖二等奖6项。发表论文100余篇，翻译出版《剑桥五重奏》（上海科技出版社2003.获得牛顿科技普及二等奖），编著教材《数据与知识工程导论》（清华大学出版社，2003）及《人工智能系统原理与设计》（国防科技大学出版社，1989）。获得与本专著有关的国家专利一项（2000）。

书籍目录

上篇

第一章 第一后继字符有序的互关联后继树索引模型3

1.1 全文检索模型综述

1.1.1 位图 (Bitmap)

1.1.2 署名文件 (Signature Files)

1.1.3 倒排表 (Inverted Files)

1.1.4 Pat树和Pat数组

1.1.5 2相邻矩阵模型

1.1.6 全文索引模型的评价标准

1.2 第一后继字符有序的互关联后继树

1.2.1 基本定义

1.2.2 后继区间概念介绍

1.2.3 创建第一后继有序的互关联后继树创建算法

1.3 后继区间查询算法

1.4 后继区间查询算法复杂度及其性能分析

1.5 实验与分析

1.6 小结

第二章 双排序互关联后继树创建与查询算法

2.1 引言

2.2 双有序互关联后继树索引创建算法

2.3 双排序互关联后继树查询算法

2.3.1 逆向区间二分查询算法

2.3.2 双排序互关联后继树二分验证查询算法

2.3.3 双排序互关联后继树线性优化查询算法

2.4 实验与分析

第三章 互关联后继树索引的编码优化方法

3.1 引言

3.2 编码方案

3.3 位编码算法

3.4 原文生成算法

3.4.1 字符定位算法

3.4.2 后继树编码计数算法

3.4.3 后继树编码值算法

3.4.4 原文根地址算法

3.4.5 原文生成算法

3.5 全文检索算法

3.6 实验数据及分析

3.7 小结

第四章 基于互关联后继树索引的文本压缩

4.1 文本数据压缩的常用技术

4.1.1 数据压缩

4.1.2 文本压缩技术的分类

4.1.3 几种主要的压缩模型

4.1.4 文本压缩技术的应用

4.1.5 压缩的评判标准

## <<数据索引与数据组织模型及其应用>>

### 4.2 后继树静态词典压缩

#### 4.2.1 互关联后继树静态词典的设计

#### 4.2.2 压缩和解压算法

#### 4.2.3 静态词典压缩算法改进

#### 4.2.4 性能比较与分析

### 4.3 互关联后继树自适应词典压缩

#### 4.3.1 互关联后继树自适应词典的设计

#### 4.3.2 压缩和解压算法

#### 4.3.3 压缩算法改进

#### 4.3.4 互关联后继树自适应压缩算法特点

#### 4.3.5 性能比较与分析

#### 4.3.6 小结

## 第五章 基于后继模式树的XML索引模型

### 5.1 引言

### 5.2 基于后继模式树的倒向XML索引

### 5.3 XML的统一索引模型

#### 5.3.1 联合索引的创建

#### 5.3.2 XML数据与全文数据的协同查询

### 5.4 XPath的自顶向下与自底向上查询

#### 5.4.1 绝对位置路径的查询树解析

#### 5.4.2 自顶向下查询

#### 5.4.3 自底向上查询

### 5.5 基于后继模式树的协同查询

#### 5.5.1 后继模式树上的路径查询

#### 5.5.2 基于后继模式树的自底向上协同查询

### 5.6 系统实现与实验

### 5.7 小结

## 第六章 基于互关联后继模型的搜索引擎

### 6.1 引言

#### 6.1.1 搜索引擎的原理

#### 6.1.2 主流搜索引擎介绍

#### 6.1.3 黄页搜索引擎基本需求

### 6.2 基于互关联后继索引的搜索引擎

#### 6.2.1 搜索引擎与互关联后继树的结合

#### 6.2.2 互关联后继树搜索引擎的索引结构

### 6.3 匹配度计算

#### 6.3.1 匹配度定义

#### 6.3.2 匹配度计算公式

#### 6.3.3 匹配度计算实现技术

#### 6.3.4 词位置号的保存

#### 6.3.5 匹配度计算

#### 6.3.6 实验与分析

### 6.4 搜索结果排序技术

#### 6.4.1 通用排序算法介绍

#### 6.4.2 基于动态划分的多权值快速排序

#### 6.4.3 基于区间的划分算法

### 6.5 小结

下篇

第七章 序列文本索引的粒子模型

7.1 引言

7.2 文本索引的粒子模型

7.2.1 序列对象有序化

7.2.2 有序化的序列对象粒子化

7.2.3 有序化的序列对象粒子的结构关系

7.3 互关联后继索引—文本序列商空间 $I_{str1, 2}$ 的性质

7.3.1 商空间的熵的性质

7.3.2 保假性与保序性

7.3.3  $I_{str1, 2}$ 的特殊性质

7.4 小结

第八章 创建索引模型的数学方法

8.1 创建全文索引模型的数学变换

8.2 互关联后继索引模型性能分析与比较

8.2.1 倒排表和Pat数组的性能分析

8.2.2 互关联后继索引模型性质

8.2.3 分析与比较

8.3 存储模型比较分析

8.3.1 原文和索引都在内存

8.3.2 原文在外存索引放置于内存的情况

8.3.3 原文和索引都放置于外存的情况

8.4 与Pat树等其他索引模型的关系

8.5 小结

第九章 互关联后继索引模型的熵与压缩原理

9.1 引言

9.2 粒子细分的方法不能降低信息量

9.3 公因子方法压缩原理

9.4 差异熵压缩的原理

9.4.1 一元编码

9.4.2 Golomb方法

9.4.3 编码模式方法和实例

9.5 小结

第十章 事务库的组织与数据挖掘

10.1 FP-Growth方法简介

10.2 隐式互关联间接后继树/图的挖掘方法

10.2.1 隐式互关联间接后继树( /图) 表示

10.2.2 ISTR+树创建算法

10.2.3 Istr+树频繁项集挖掘算法

10.2.4 Istr+树挖掘算法与FP-growth算法的比较

10.3 可变维数的隐式间接互关联后继树的挖掘方法

10.3.1 可变维数的隐式间接互关联后继树表达

10.3.2 T-Istr+间接后继表的性质与频繁项的挖掘算法

10.3.3 算法复杂性分析

10.4 小结

第十一章 关系数据库与演绎数据库的数据组织

11.1 协同查询问题的回顾

## <<数据索引与数据组织模型及其应用>>

- 11.2 关系的互关联后继的数据表达
- 11.3 关系R-Istr+互关联隐式间接后继索引表与演算
  - 11.3.1 基于R-Istr+索引表的关系演算
  - 11.3.2 R-Istr+查询操作的复杂性分析
- 11.4 基于R-Istr+的关系库的协同查询
- 11.5 演绎数据库的索引与演绎
  - 11.5.1 演绎数据库的索引
  - 11.5.2 基于I-Istr+的基本查询算法（集合查询算法）
  - 11.5.3 演绎数据库的演算
  - 11.5.4 对规则的演算
  - 11.5.5 复杂性的对比
- 11.6 小结

### 第十二章 逻辑程序或知识库的索引

- 12.1 逻辑程序的简单介绍 [ Hu88 ]
- 12.2 严格有序的逻辑程序的索引模型
  - 12.2.1 第一种函词的序列表示
  - 12.2.2 创建逻辑子句索引的步骤
  - 12.2.3 互关联后继树索引与Warren抽象机数据的关系
  - 12.2.4 基于互关联后继索引的合一操作——索引合一
  - 12.2.5 推理步骤说明
  - 12.2.6 子句的或并行
  - 12.2.7 串行执行机制的改进——子目标级别优选
- 12.3 逻辑程序并行性
  - 12.3.1 第二种函词表示方法
  - 12.3.2 创建互关联后继索引的步骤
  - 12.3.3 合一中的并行
  - 12.3.4 逻辑程序并行推理的实例
- 12.4 结论

### 第十三章 基于互关联后继数据组织模型的文本信息隐藏技术

- 13.1 引言
- 13.2 相关技术与术语介绍
  - 13.2.1 术语介绍
  - 13.2.2 相关技术介绍
- 13.3 基于互关联后继索引模型的文本信息隐藏方法
  - 13.3.1 信息隐藏过程
  - 13.3.2 信息提取过程
  - 13.3.3 控制功能
- 13.4 安全强度分析
- 13.5 小结

### 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>