

<<XML数据查询与信息检索系统>>

图书基本信息

书名：<<XML数据查询与信息检索系统>>

13位ISBN编号：9787508471518

10位ISBN编号：7508471512

出版时间：2010-3

出版时间：水利水电出版社

作者：陈忠明

页数：121

字数：168000

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

XML已经成为互联网上数据表示和数据交换的标准。

随着XML文档数据量和文档数量的快速增长,产生了很多问题,其中很重要的一个问题就是如何有效地查询这些文档,也就是结构化查询,亦称为数据查询。

而有效的数据查询又涉及文档的存储机制和索引结构等问题,这些问题已经引起了学术界和工业界广泛的研究热情,学者们在这些问题上作了大量的研究工作。

另外一个问题是基于XML。

文档的信息检索,这也是一个新产生的研究问题。

面对巨大的网络信息,如何才能为用户检索出真正有效的信息是一个非常具有挑战性的研究问题。

现存的大部分搜索引擎是基于关键字搜索的,页面排序算法采用页面的超链接或页面内容的文本特性

。如果页面采用XML,来编写,那么就需要合理地利用XML的结构特性、语义特性以及其他的相关性质来提高检索的效果和效率。

这就促使对XML文档进行信息检索成为了一个非常有意义的问题。

XML上的信息检索正开始受到学术界的高度重视。

对XML进行结构化查询和信息检索是两个既具有相关性又具有不同特性的研究问题,如何集成这两个研究问题就更加具有研究意义。

本书的主要研究对象就是如何改进XML的结构化查询以及信息检索的相关理论与技术,以便于它们更好地集成在一起,从而可以更加优化地执行用户的查询需求。

针对这个研究目标,本书做了大量的研究工作。

首先,本书在XQuery语言的基础上扩充了全文本检索功能,为了与原来的XQuery区分,扩充后的语言称为XQuery+(XQueryPlus)。

XQuery+语言有如下特点:在XQuery+语言里,扩充了XQuery的检索功能,增加了一个为检索服务的谓词;在XQuery+中还支持检索词的布尔操作。

本书的主要研究任务之一是如何有效地处理XML的结构化查询。

作为处理XML结构化查询的基础,XML文档的节点编码模式和索引结构是研究的核心问题。

本书提出了一个新颖有效的节点编号模式,详细地讨论了节点编号模式的定义和性质。

节点编号模式为XML文档索引和查询提供了基础,一个有效的节点编号模式应该可以包含结构信息,易于支持索引和查询。

从本书给出的节点编号定义和性质分析,我们知道编号模式可以满足这些基本的要求。

本书在这个节点编号模式的基础上建立了一个HiD索引结构,HiD索引结构有效地集成了结构索引和值索引两个部分。

通过大量有竞争性的实验分析表明,采用HiD索引机制方法可以在索引的构建时间和空间消耗上得到很好的平衡和性能表现。

## <<XML数据查询与信息检索系统>>

### 内容概要

本书主要研究改进XML数据查询和信息检索的相关理论与技术，以便于它们更好地集成在一起，从而可以更加优化地执行用户的查询需求，针对这个研究目标，本书做了大量的研究工作。

本书提出了一个新颖有效的节点编号模式，详细讨论了节点编号模式的定义和性质，还提出了一种新颖有效的对基于XML信息检索查询进行相关度打分的算法，这个打分机制结合了检索查询关键词的频度、文档的结构化特性、文档的语义特性等。

基于对结构化查询和信息检索的相关研究成果，本书提出了处理XML结构化查询和信息检索的有效算法与机制。

本书还讨论了一个原型系统的设计目标、分析与设计过程。

## &lt;&lt;XML数据查询与信息检索系统&gt;&gt;

## 书籍目录

前言第1章 绪论 1.1 研究背景 1.2 XML介绍 1.2.1 元素(Element) 1.2.2 属性 1.2.3 指令/处理指令 1.2.4 注释 1.2.5 CDATA 1.2.6 XML的语法规则 1.3 Xpath介绍 1.3.1 节点(Node) 1.3.2 XPath谓词 1.3.3 XPath轴 1.3.4 XPath节点测试 1.4 XQuery介绍 1.4.1 XQuery的语法 1.4.2 XQuery的运算符 1.4.3 XQuery函数 1.4.4 XQuery条件表达式 1.5 相关研究 1.5.1 数据库的研究 1.5.2 XML数据管理 1.5.3 XML数据查询 1.5.4 信息检索及基于XML的信息检索 1.6 小结第2章 基于XQuery的信息检索语言 2.1 XML查询语言 2.2 XML信息检索语言的特点 2.3 XML信息检索语言XQuery+ 2.3.1 XQuery+语法分析 2.3.2 XQuery+语义分析 2.4 小结与问题第3章 XML节点编号模式与索引结构 3.1 预备知识 3.2 XML节点编号模式 3.2.1 节点标签路径数 3.2.2 节点数据路径数 3.2.3 节点标识 3.3 XML索引结构 3.4 值索引结构 3.5 相关实验及分析 3.6 小结与问题第4章 XML相关度打分机制与算法 4.1 问题描述 4.2 IR查询表达 4.3 相关度打分机制 4.3.1 结构相关度 4.3.2 语义相关度 4.3.3 相关度集成 4.4 实例分析 4.5 实验与分析 4.6 小结与问题第5章 查询处理 5.1 问题描述 5.2 XQuery查询处理算法 5.2.1 单路径查询 5.2.2 树模式查询算法 5.3 XQuery+查询处理算法 5.4 XQuery查询实验分析 5.5 XQuery+查询算法实验分析 5.6 小结与问题第6章 原型系统的设计与实现 6.1 原型系统分析与设计 6.1.1 系统设计目标和原则 6.1.2 需求分析 6.1.3 数据流图 6.1.4 系统架构 6.2 原型系统模块分析 6.2.1 系统处理流程 6.2.2 模块设计 6.3 原型系统的实现 6.3.1 原型系统核心数据结构 6.3.2 查询引擎处理过程部分代码分析 6.3.3 原型系统界面及使用介绍 6.4 小结与问题第7章 结论与展望参考文献

## 章节摘录

插图：查询、存取等。

信息检索是一个一直以来都在研究的热点问题，有大量研究者从事信息检索的研究。

XML的结构化查询主要是基于XML的结构上精确的数据查询，当用户给出一个：XML的查询时，要求查询系统返回的是精确的查询结果节点集；而XML，的信息检索主要是基于XML文档的内容，即以文本信息为主的。

用户给出的一般是文本信息的要求，而结构信息等可能是模糊的、不精确的，要求返回的节点可能在查询时也不会被清晰地定义好。

信息检索的结果是给出和用户要求的文本信息最相关的节点集或者最相关的前K个结果，也就是所谓的TOP-K查询。

在信息检索领域，涌现了大量新型智能检索技术，如职能检索、知识挖掘、全息检索等。

下面简单介绍一下这些技术。

#### (1) 智能检索。

智能检索利用分词词典、同义词典、同音词典改善检索的效果。

比如用户查询“计算机”，那么与“电脑”相关的信息也能检索出来；进一步还可在知识层面上辅助查询，通过主题词典、上下文词典、相关同级词典形成一个知识体系，给予用户智能知识提示，最终帮助用户获得最佳的检索效果。

智能检索还包括歧义信息检索处理，如“苹果”究竟是指水果还是指电脑品牌，将通过歧义知识描述库、全文索引、用户检索上下文分析以及用户相关性反馈等技术结合处理，高效、准确地反馈给用户，使其得到最需要的信息。

#### (2) 知识挖掘。

知识挖掘主要指文本挖掘技术，目的是帮助人们更好地发现、组织、表示信息，提取知识，满足信息检索的高层次需要。

包括摘要、分类（聚类）和相似性检索等。

自动摘要就是利用计算机自动地从原始文献中提取文摘。

在信息检索和服务中，自动摘要有助于用户快速评价检索结果的相关程度和多种形式的内容分发。

自动分类可基于统计或规则，经过机器学习形成预定义分类树，再根据文档内容特征将其归类。

自动聚类则是根据文档内容的相关程度进行分组归并。

相似性检索技术基于文档内容特征检索与其相似或相关的文档，是实现用户个性化相关反馈的基础，也可用于去重分析。

#### (3) 异构信息整合检索和全息检索。

在信息检索分布化和网络化的趋势下，对于信息检索系统的开放性和集成性要求越来越高，需要它能够检索和整合不同来源和结构的信息，这是异构信息检索技术发展的基点，包括支持各种格式化文件的处理和检索；支持多语种信息检索；支持结构化、半结构化及非结构化数据的统一处理等。

全息检索即支持一切格式和方式的检索，从目前实践来看，已经发展到异构信息整合检索的层面，而基于自然语言理解的人机交互以及多媒体信息检索整合等方面尚有待进一步突破。

从实际的应用角度分析，信息的充分利用和交流是实现真正的信息化的基础。

美国财富500强企业网站信息检索工具拥有率几乎达到了100%，而中国企业的信息利用状况则形成了极大的反差。

企业只是追逐时尚，花费大量的人力物力构建自己的网站和丰富的内容体系，然而信息检索工具却不尽人意。

因此，发展信息检索方面的应用是当前中国企业信息化建设的重要任务。

## <<XML数据查询与信息检索系统>>

### 编辑推荐

《XML数据查询与信息检索系统》是由中国水利水电出版社出版的。

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>