

XML索引与查询的若干关键技术分析*

蔡劲松

(安徽新闻出版职业技术学院, 安徽 合肥 230601)

【摘要】信息科技时代计算机用户数量持续增加,原有计算机设备在处理数据资源时承受的载流量越来越大,这要求计算机系统具有更高水平的数据处理能力。可扩展标记语言是计算机数据自动化处理的必然趋势,采用更加高端数字语言执行数据库操控命令,维持了数据资源调控利用的最优化配置。本文分析了可扩展标记语言的功能特点,对可扩展标记语言索引与查询常用关键技术展开分析,编制了一套先进的数据标记与定义方案。

【关键词】可扩展标记语言;索引;查询;关键技术

【中图分类号】TP311.13 **【文献标识码】**A **【文章编号】**1673-1891(2014)04-0078-03

计算机应用技术已经对社会生活产生了深远影响,凭借其强大的数据处理能力,在社会各个行业中得到推广应用。近年来,无论是企业用户或个人用户,社会信息活动对计算机依赖程度越来越大,这就要求数据库系统具备更加优越的数据条件,才能更好地执行数据操控流程。可扩展标记语言(XML)索引与查询是数据库运行的先进技术,用其解决传统数分配存在的缺陷,为用户建立更加综合性的控制平台。

1 XML应用特点

为了改变传统数据库模式的不足,新时期研发了可扩展标记语言层次,这种数据模式可以提升原有数据库的工作效率,解决单一数据传输平台存在的功能问题。从实际应用来看,XML应用特点包括:可扩展、自定义、结构化等特点。

(1)可扩展。XML采用了一种可扩展标记语言,对原始数据集中化处理后形成稳定的层次,加快了新型数据结构的最优化改进,这是保障原始数据状态的关键。XML本质上也是标准通用标记语言的子集,一种用于标记电子文件使其具有结构性的标记语言,这就说明了XML数据库在任何时候都可以扩展,按照用户使用要求编排新的指令结构^[1]。未来可扩展文焕数量不断增加,XML将成为计算机数据处理的主流技术。

(2)自定义。计算机用户涉及到多个行业领域,每一位用户所用数据资源形式不一,这些都会影响到最终数据层的执行效率。XML技术在可扩展性基础上,为用户提供了自定义功能,根据个人使用要求对某个电子文焕定义、标记,改进了旧式电子文焕的可利用功能。例如,XML可以用来标记数据、定义数据类型,帮助用户快速查找、筛选、执

行数据命令,短时间内为用户提供优越的数据服务,按照规定标准命名各类数据文件。

(3)结构化。数据资源在使用阶段要进行传输处理,XML可建立相对稳定的链路传输环境,避免数据形式单一化带来的不良影响。当前,电子文件或数据包等标记语言需通过网络传输,而XML能够最大范围满足万维网运行要求,创建符合计算机交换器平台的程序语言,解决了传统数据层次的功能缺失。结构化数据模式按照语言标准次序,完善了用户现有数据库的执行线路,从安全性与实用性等方面均可产生优越的数据保护作用。

2 基于路径的XML索引技术

按照不同路径完成XML索引与查询功能,这是基于XML树的搜索操作,可以从多个方面手机与XML相关的数据信息。如图1,筛选一种简洁式的树结构,以多种元素构建成XML索引路径,这样可以避免在同一路径上出现多个节点。早期数据库索引与查询系统中,采用了多个搜索路径方案,虽然为用户提供了更多查询方式,但普遍存在着相同路径、相同节点等问题,降低了数据资源传输的稳定性,一些重要数据在传输阶段也会丢失^[2]。

(1)综合查询。XML索引与查询系统中,根据用户实际操作指令要求,建立了更加完整的数据资源方案,领域应用多个路径模式执行查询工作。比较常见的索引技术包括:DataGuides索引、Index Fabric索引、XML自适应索引等,这些都可根据用户工作要求执行最优化索引方式。

如图1,采用XML文档建立某一种树形结构,从多个方面反映了数据库层次的分布情况,利用最基本的根元素完成动态操作。在图1中,XML内含有多个根元素,这是结构其它元素的父元素,这些元

收稿日期:2014-07-29

*基金项目:2011年省级教学质量工程项目之省级特色专业-电子出版技术(皖教高[2011]5号,编号199)。

作者简介:蔡劲松(1979-),男,安徽省无为县人,硕士,讲师,研究方向:数据库程序设计与开发。

素在XML文档结构层中得到扩展,结构树从最底部开始,逐渐扩展到最顶端,这也说明了XML索引与查询的多功能特点。此外,父、子作为元素关系的描述性语言,可根据树形结构进一步完善,加快了元素层构造功能的完整性。

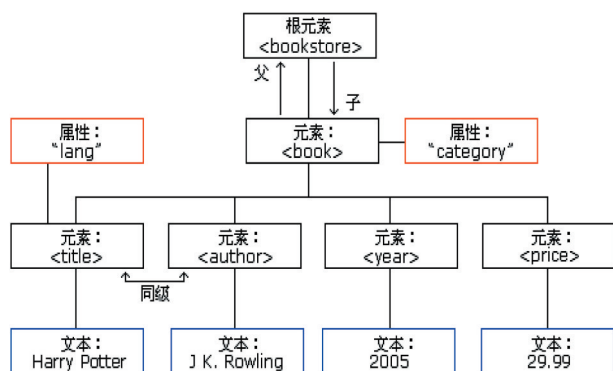


图1 XML树结构

(2)无线通信。移动互联网是第三代(3G)、第四代(4G)移动通信技术的核心支撑,借助移动网可实现数据信号的自由化传输,维持发送端、接收端等数据流程的稳定运行,加快了移动互联网结构层次的优化改进。XML索引与查询需要充分利用3G、4G技术实现自动化处理,才能满足现代远程控制系统的功能要求,不断优化数字信号的传输模式,这些都是未来远程控制系统功能改革的主流趋势。

3 基于节点的XML索引技术

XML技术平台设定了节点索引方式,这是根据电子文件组织元素创建的新查询方式,避免数据重复或二次查询等产生的不便。选择节点作为XML索引平台,实现了数据单元结构的层次化,按照不同规律编排原始数据层,提升了XML索引的操作效率。此外,基于节点的索引流程中,可随时对数据单元实施定位处理,找准与数据结构相关联的节点位置,保证数据库资源得到全面利用。可以分为以下几类:

(1)基于前缀的索引。选择某一个前缀作为搜索目标,按照文件传输路径执行各项步骤,这些都是提升XML索引效率的关键技术。新时期计算机用户对数据处理操作标准不断提升,维持高标准数据操控可促进资源优化配置。XML索引与查询系统采用前缀方法,基本上锁定了所要查询对象的位置,不会出现类似文件名同时搜索等问题。基于前缀的索引主要是根据Dewey编码生成的索引,文献的ORDPATH编码采用的也是类似的方法,并给出了压缩ORDPATH的方法,该方法已应用于SQL Server 2005的索引组织中。

(2)基于区间编码的索引。任何数据在编制过程中会设定相应的区间,结合区间编码执行索引操作,可实现XML技术功能的稳定性。按照用户操作数据的要求,XML树T中的每一个结点被赋予一个区间编码[begin,end],一个结点的区间编码包含它的后裔结点的区间编码^[3]。XML索引与查询根据语句结构的分解形式,按照特殊信号命令完成数据编排,把准确的数据结果呈现给用户。

4 基于传输的XML索引技术

面对日趋增多的数据库资源,用户如何快速索引、查询到有价值信息,这是计算机操作系统急需改良的方面。计算机网络数据传输也是XML索引与查询的有效方式,这是根据用户数据传输路径设定的关键技术之一,满足了不同传输需求用户的动态控制要求。

(1)P2P技术。互联网是数据传输的主要平台,点对点技术(P2P)作为一种常用的传输方式,如图2,用于XML索引操作具有个性化特点。这是由于P2P在索引空间、路径等方面基本确定,无需用户大范围选择传输路径,只要确定发送端、接收端等位置即可^[4]。P2P技术提升了XML索引查询的针对性,缩短了用户处理数据的操控难度。

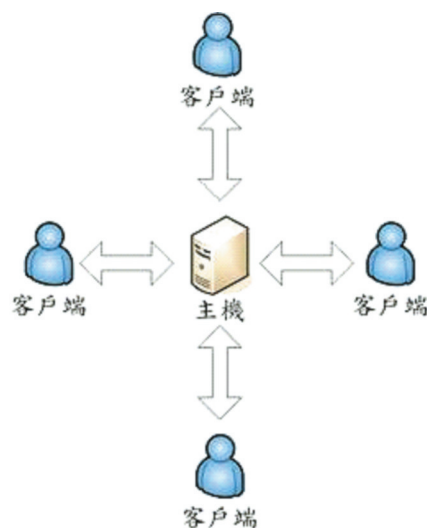


图2 P2P传输技术

(2)P2MP技术。相对于P2P技术,P2MP技术的传输范围更广,为XML提供了多信道索引服务平台,让用户查询到更加广泛的数据资源。P2MP技术可从某一个位置到多个位置进行传输,创建多个信道传输服务路径,进而适应了大范围索引、查询、收录、下载等操作要求。

5 结论

数据库是计算机应用系统的核心构成部分,配

备高性能数据库可实现资源优化配置,为用户创建更加稳定的数据使用平台。随着计算机服务群体的扩大化发展,数据库搜索、查询等操作流程更加复杂化,选择一种高效率的查询方式,可有助于数

据库层次的升级改造。XML索引与查询技术中,既可以按照路径、节点等方式查找,也可根据P2P、P2MP等通信传输平台进行操作,极大地方便了用户对数据的传输控制。

注释及参考文献:

- [1]雷向欣,胡运发,杨智应,等.基于互关联后继树的XML索引技术[J].计算机研究与发展,2005,12(7):23-25.
- [2]王静,孟小峰,王宇,等.以目标节点为导向的XML路径查询处理[J].软件学报,2005,14(5):6-8.
- [3]马海兵,张成洪,张锦,等.基于IS~±树模型的频繁模式挖掘[J].计算机研究与发展,2005,14(4):33-34.
- [4]万常选,刘云生,徐升华,等.基于区间编码的XML索引结构的有效结构连接[J].计算机学报,2005,16(10):9-11.

Analysis of Some Key Technologies of XML Index and Query

CAI Jing-song

(Anhui Vocational College of Press and Publishing, Hefei, Anhui 230601)

Abstract: With computer users increasing in the era of information technology, the original computer equipment's beard load flow in processing data resource is more and more big, which requires computer system with higher levels of data processing ability. The extensible markup language is the inevitable trend of the automation of computer data processing, using more high-end digital language perform database control command, to maintain the optimal allocation of resources utilization data regulation. This paper analyses the feature and function of extensible markup language, common key technologies of extensible markup language index and query, develops a set of advanced data markers and defined scheme.

Key words: extensible markup language; index; query; key technology