

面向内容管理的浏览器设计与实现

庄军, 朱斐, 张芸

(苏州大学 计算机学院, 江苏 苏州 215006)

摘要:分析了传统文件系统在企业信息管理中的不足,提出了面向内容管理的必要性。给出了面向内容管理浏览器的设计目标和功能。介绍了文件在面向内容管理的系统中是如何组织和管理的,并基于 Windows 平台演示了系统的部分功能。文中最后对面向内容管理的浏览器做出了客观性评价。文中提出的面向内容管理的浏览器能够充分挖掘信息的价值,有助于用户更好地进行内容管理,具有较好的研究价值和实用意义,有利于促进内容管理的发展。此外,文中对支持内容管理的文件管理器实现技术所进行的探讨,对于该领域的开发者应具有良好的借鉴意义。

关键词:内容管理;元数据;IFS;文件管理

中图分类号:TP393.092

文献标识码:A

文章编号:1673-629X(2006)12-0175-03

Design and Implementation of a Content Management - Based Explorer

ZHUANG Jun, ZHU Fei, ZHANG Yun

(Computer School of Soochow University, Suzhou 215006, China)

Abstract: Analyzes the disadvantages of enterprise management system based on traditional file system, and brings forward a system based on content management. Then make clear the destination and function of our system. After that describe how files are organized and managed in detail and demonstrate the functions of the Explorer on the Windows platform. In the last part of this paper, evaluate our system on objectivity. The Explorer based on content management designed is so useful that the users can do well in content management. And the value of information can be well mined. So system is helpful to promote the development of content management.

Key words: content management; metadata; IFS; file management

0 引言

互联网技术和电子商务应用的发展迅速地改变着企业的运作和经营管理模式,其中影响最大的将是企业的信息存储系统。Forrester Research 的一个调查显示:企业信息的内容量在以每年 200% 的速度增长。目前,传统的文件系统按照文件的名称来访问数据。在海量的数据下的环境下,传统的文件系统所提供的信息访问方式效率低下,提供的文件操作过于简单,文件的管理没有面向文件的内容,分布式文件的信息的协同工作非常困难等。急需一种新的、面向内容管理的文件系统来解决这些问题。

Oracle 公司在其核心数据库上支持了内容管理,提供了基于内容管理的开发平台 IFS^[1],2004 年 12 月 9 日 Oracle 公司发布企业级内容管理工具——Oracle Files 10g^[2],并宣布今后将跨越数据库加大内容管理领域的投入。

微软也在内容管理领域投入了大量的精力,为用户提供了一些内容管理功能。Windows 的下一代系统 Longhorn 就是面向内容管理的。Longhorn 将改用基于

SQL 的文件管理模式,提供基于内容检索的模糊查找功能^[3]。

1 系统设计

面向内容管理的文件系统可以以传统的文件系统作为内容的存储库,这种体系只适用于松散控制环境下较小的情况。另一种是以数据库作为内容的存储库,这种体系是建立在数据库管理系统之上,能够提供更多的功能和稳定的性能。

针对目前文件系统的不足,提出了一种面向内容管理的文件系统,并且选择使用数据库作为内容的存储库,将文件的内容和其各项属性都存放在数据库中。把管理体系建立在数据库管理系统之上,能够充分利用数据库自带的强大的管理和搜索功能,实现分布式存储和访问的高效性^[4]。另外,建立在面向对象的数据库的基础上的文件系统,具有了面向对象的文件系统的特征。

1.1 面向内容管理的浏览器

面向内容管理的浏览器提供强大的文件管理功能。分散的企业用户可以把需要管理的文件上传到数据库中,设置和查看文件的相关属性,并为文件设定一个访问权限。用户可以将文件版本化^[5],保存文件的不同版本,检出版本化文件,进行查看和修改,作为一个新版本存入数

收稿日期:2006-03-20

作者简介:庄军(1985-),男,江苏常州人,研究方向为内容管理、管理信息系统。

据库。

基于内容管理的文件存储在数据库中,这样可以利用数据库的搜索功能提供强大的内容管理功能。除了支持传统的对文件名称、创建时间、类型的搜索外,还将支持更有价值的搜索方式,如全文搜索、类别搜索和关联搜索等,同时还可以指定搜索结果的对象类型、是否包含子对象、是否包含版本化文件等。面向内容管理浏览器的系统功能框架如图 1 所示。

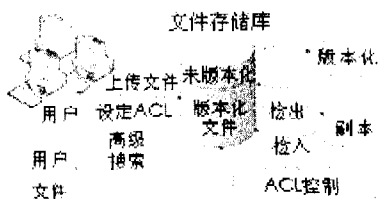


图 1 系统功能框架图

1.2 扩展的文件管理

文件的属性可分为四种相对独立的信息:普通属性信息、类别信息、版本信息、安全控制信息。下面分别介绍各种信息在面向内容管理系统中是如何管理的。

1.2.1 文件元数据的管理

元数据(Meta-data)是“关于数据的结构化数据”,用于描述数据的内容、质量、表示方式、空间参考、管理方式以及数据集的其他特征。文件的普通属性信息和类别信息都是文件的元数据^[6,7]。

文件普通属性信息和类别信息的管理比较简单。文件的普通属性很容易获得。在获得类别属性时,先要判断文件是否版本化。对于未版本化的文件,可以直接通过类别类型数组获得。对于已版本化的文件,先获取文件的最新版本或者指定版本,再获取类别类型数组,从类别类型数组中再逐个获取类别的属性。这些信息都位于数据库表中,有一些属性如果被修改了就会引起紊乱或不一致,如文件的创建日期、创建者等,因此这些属性在文件创建后就不能被修改。对于已经版本化的文件,只能修改最新版本的相关属性值。文件历史版本的内容和属性值都是只读的,用户不能对它们进行修改。

1.2.2 文件的版本控制

文件的版本控制是内容管理的一个重要功能。通过版本操作,用户保存同一文件的不同版本,从而保存创作的历史轨迹。在需要的时候,用户可以检出历史版本,对其进行修改。同时,通过检入/检出功能可以实现多用户的协同编辑。文件的版本化就是为此文件创建三个管理文件多个版本的对象(Family, Version Series, Version Description),当通过检入/检出为版本化文件创建一个新的版本操作时,实际上就是为文件创建两个新的对象(Version Description, Document)。文件版本化的过程如图 2 所示。

当用户检出文件时,先使用一个副本保存文件内容。用户修改完成后,通过检入操作把它作为最新的内容保存

到新版本中。

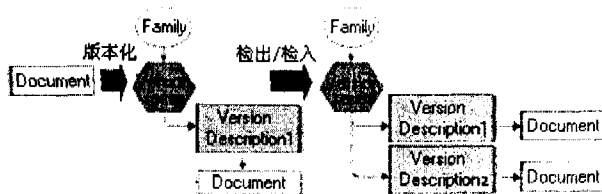


图 2 文件版本化过程

1.2.3 文件的安全控制

基于内容管理的文件系统通过双重的安全控制来保证系统的安全性^[8]。

首先当用户试图访问信息库中的信息时,必须提供登陆名和口令。只有合法的用户才能建立与信息库的连接,访问信息库。用户通过了信息库层的访问控制验证,访问信息库中文件时,还必须通过 ACL(Access Control List)验证。

ACL 实现了对文件更细粒度的安全访问控制。ACL 是由一个或多个 ACE(Access Control Entry)组成,每个 ACE 由被授权人和与之对应的一组许可权限组成^[9]。

在基于内容管理的文件系统中,每个文件都拥有一个 ACL,在文件被创建时,系统会为其自动分配一个默认 ACL,授权用户可以查看/修改文件的 ACL,并允许创建新的 ACL。在修改文件 ACL 时,首先要从系统中取出所有能够供用户使用的 ACL,当没有满足用户要求的 ACL 时,用户可以创建一个新的 ACL,并将其赋给文件,修改文件的 ACL 时,首先要根据已选择的 ACL 名称来获取 ACL 对象,然后重新设置文件的 ACL。对于文件夹的修改,则要判断是否修改文件夹的 ACL,还是包括其子文件。但如果要修改文件夹中所有文件的 ACL,则需要通过递归调用的方法完成修改动作。

1.3 系统框架

系统包括了客户端模块、文件管理模块和文件存储模块三个部分。

客户端模块通过可视化的方式提供给用户使用,实现了用户的操作逻辑。用户发送的文件访问请求,通过调用挂载到本地文件系统的接口,传送到文件管理模块。文件管理模块接受到操作命令后,根据操作命令进行相应的逻辑处理,执行文件的访问权限方案,从文件存储库中调用文件,进行逻辑处理后再把结果传递到客户端。最后,客户端模块将结果显示给用户。系统框架结构如图 3 所示。

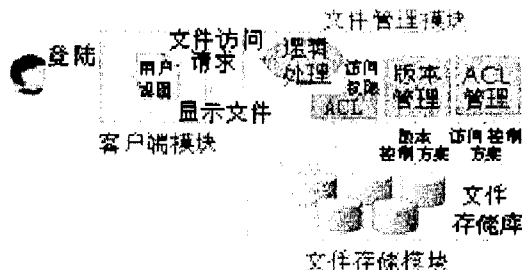


图 3 系统框架结构图

2 系统实现

Oracle 9iFS 是 Oracle 公司开发的基于 Oracle 数据库的新型文件系统。Oracle 9iFS 将企业所有数据统一成单个、统一的信息库,企业的所有数据都存放在 Oracle 数据库中。Oracle 9iFS 被设计成数据库层、中间层和客户端三层架构来提供较好的性能、可扩展性和可靠性,它已经为文件管理提供了一系列的 Java API 开发包,所以笔者设计的面向内容管理的浏览器是基于 9iFS 开发的。

面向内容管理的浏览器的用户界面被设计成类似于传统文件系统的风格,用户可以在客户端进行简单的逻辑操作。

2.1 右键菜单

为了用户操作的方便,将对文件的操作加入到右键菜单中,每项菜单命令对应一个文件操作。用户可以通过右键菜单,调用系统的各项功能。如图 4 所示。

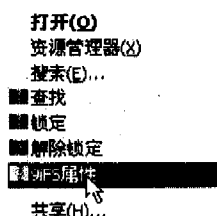


图 4 右键菜单

2.2 高级搜索页面

面向内容管理的浏览器除了提供传统文件系统提供的搜索功能外,还提供了全文搜索、正则搜索、类别搜索和关联搜索等高级搜索。不同类型的搜索分布在不同的 Tab 页中,用户可以根据提供的搜索功能进行相应的查询。在一次搜索中,用户还可以进行组合条件搜索。

高级搜索如图 5 所示。

名称和位置	日期和大小	高级	内容	类别	关联查找
对象类型: DOCUMENT					
关联属性: CONTENTOBJECT					
关联对象: CONTENTOBJECT					
设置关联对象查找条件:					

图 5 高级搜索

全文搜索时,在查找操作界面的“内容”Tab 页中输入全文搜索的关键字,然后执行搜索,就可以将内容包含关键字的所有文件搜索出来。

在“名称和位置”Tab 页中为高级用户提供了正则搜索功能,用户可以对字符串类型的属性构建复杂的正则查询条件,准确地搜索到目标文件。

在面向内容管理的浏览器中,可以通过类别属性对文件进行分类,这样存放在不同位置的文件可以通过类别将同类文件搜索出来。

关联搜索能够为用户进一步提供搜索功能,如某用户需要根据邮件发送者的一些个人信息(如:爱好、生日)来搜索满足条件的邮件。

3 小结

本系统是针对用户对内容管理的迫切要求和当前内容管理系统的不完善而设计的,为用户提供了强大的文件管理功能,提高了文件访问的安全性和效率,实现了跨平台性,并将所有的功能操作集成到了右键菜单中。面向内容管理的浏览器实现了基本的内容管理功能,体现了内容管理的优越性,具有良好的研究价值和实用意义。

参考文献:

- [1] Hyman F, Garg S, Harrison S. Oracle® Internet File System Installation Guide Release 9.0.1.1.0 for Microsoft Windows NT/2000[EB/OL]. 2001-09. <http://www.oracle.com>.
- [2] Stokes A, Dawson D. Oracle® Internet File System Developer Reference Release 9.0.1.1.0 Oracle 9iFS and Oracle9i database documentation sets[EB/OL]. 2001-09. <http://www.oracle.com>.
- [3] Rizzo T. 新型 Windows 文件系统简介[EB/OL]. 2004-08-25. <http://www.microsoft.com/china/MSDN/library/windev/longhorn/winfs03112004.msp>.
- [4] 车敦仁,周立柱,王令赤. 面向对象数据库系统的体系结构[J]. 软件学报,1995(10):599-606.
- [5] 周 军. 应用版本控制软件管理软件开发[J]. 计算机系统应用,2000(10):50-52.
- [6] 潘 定,沈钧毅. 数据仓库中实时元数据管理的研究[J]. 计算机工程,2000(5):29-31.
- [7] 王 强,刘东波,王建新. 数据仓库元数据标准研究[J]. 计算机工程,2002(12):123-125.
- [8] 朱 斐. 一种结构化文件的访问控制模型的设计和实现[J]. 微机发展,2005,15(4):132-134.
- [9] 诸 晔. 用 ACL 实现系统的安全访问控制[J]. 计算机应用与软件,2005(3):111-114.

(上接第 174 页)

- [8] Han Jiawei. 数据挖掘概念与技术[M]. 范 明,孟小峰等译. 机械工业出版社,2001:14-17.
- [9] 何 明,李 博. 粗糙集理论框架下的神经网络建模研究及应用[J]. 控制与决策,2005(7):65-68.
- [10] 尹 松,周永权,李陶深. 基于稀疏差异度的聚类方法在信息分类中的应用[J]. 计算机技术与发展,2006,16(1):117-119.

- [11] 李宝林,王秀峰. CBRDI:一种基于范例推理的数据集成方法[J]. 计算机工程与应用,2003(16):52-55.
- [12] Beckmann N, Kriegel H P, Schneider R, et al. The R*-tree: An Efficient and Robust Access Method for Points and Rectangles[C]//Proc. ACM SIGMOD Int. Conf. on Management of Data. Atlantic City, NJ:ACMPress,1990:322-331.