

Web 文档管理系统与 NAS 结合的实现

苗元亮,滕至阳,卢 飞

(东南大学 计算机科学与工程学院,江苏 南京 210096)

摘 要:随着企业文件数据量的飞速增长,文档的存储与管理对企业提出了更高的要求。传统的在 NAS 上通过共享目录管理文档的方式变得困难,为了提高企业的办公效率,文档管理系统与 NAS 结合成为一种需求。为了满足这种需求,提出一套将 Web 文档管理系统与 NAS 管理系统结合的解决方案。通过引入一种基于 Linux 环境的开源文档管理系统,对其进行分析,并做了一些改进,以更方便实际使用,最后成功将该 Web 文档管理系统移植到了 NAS 存储管理平台,结果表明二者结合是有效的。

关键词:NAS;文档管理;移植;多语言

中图分类号:TP309.3

文献标识码:A

文章编号:1673-629X(2010)02-0129-04

Implementation of Web - Based Document Management System and NAS Combination

MIAO Yuan-liang, TENG Zhi-yang, LU Fei

(School of Computer Science and Engineering, Southeast University, Nanjing 210096, China)

Abstract: With the rapid growth of enterprises document data, the storage and management of document have put forward higher request to enterprises. Document management through sharing directories with NAS turns difficult in the traditional way. In order to improve office efficiency, the combination of web - based document management system and NAS is in demand. Propose a solution for the combination of web - based document management system and NAS management system to meet the demand. By introducing an open source document management system based on Linux and through the analysis of it, make some improvements to make it more convenient to use and successfully transplant this web - based document management system into NAS management platform. Results show that the combination is effective.

Key words: NAS; document management; transplantation; multi - language

0 引 言

NAS (Network Attached Storage, 网络附加存储)^[1,2]是一种专业网络数据存储备份设备,面对企业飞速增长的数据量, NAS 方便连接、适宜的价格、更高的容量及 RAID 保护,满足了中小企业既要保证数据安全又要降低存储成本的需求,成了中小企业数据存储与数据保护的主要角色。

NAS 利用 FTP、NFS、CIFS、AFP 等实现跨平台的文件共享,作为企业文件服务器, NAS 在文件管理方面往往有以下缺点:

- (1)不支持文件搜索;
- (2)权限管理不细致,无法针对文件进行授权,多

层次文件授权难以实现;

(3)用户间协作性差,仅通过简单的共享实现交互,没有编辑控制、消息订阅、文件工作流等功能;

(4)文件历史操作记录难以查看和分析,无法实现监控,等等。

对于以上问题,为了提高企业办公效率,需要对 NAS 文件管理功能进行改进。考虑两种改进方法,一种是另外添加专用文件服务器, NAS 主要负责文件的存储备份。这种组合可以提高文档管理效率,但是文件服务器的硬件、软件、管理维护成本,以及系统安装使用复杂性往往都超过了 NAS 自身,这也是许多企业用 NAS 单独作为文件服务器的主要原因。为此,文中提出第二种方法,将文档管理系统集成^[3]到 NAS 上。

文档管理系统可以为企业提供一个安全、可靠、开放、高效的文档管理功能,提高企业的办公效率和文件管理的综合水平。基于 Web 的文档管理系统具有良好的移植性和扩展性。实现了 Web 文档管理系统(以下

收稿日期:2009-06-21;修回日期:2009-09-12

作者简介:苗元亮(1985-),男,硕士研究生,研究方向为 Linux 软件开发与网络存储;滕至阳,教授,研究方向为操作系统、存储网络与知识工程。

简称 Web DMS)^[4,5]与 NAS 管理系统的融合。对于许多部署 NAS 的中小企业来说,将文档管理系统融合到 NAS 上,不仅可以减小企业文档管理的复杂度和软件部署的工作量,还可以充分利用 NAS 的存储备份优势,提高文档管理系统自身性能。

1 总体设计方案

NAS 往往通过专用的管理软件来配置管理。典型的 NAS 管理软件可以自动检测网络中的存储资源、配置管理 RAID、备份恢复数据、监视存储的利用率等。通过 Web 浏览器访问 NAS 管理系统,即可在上面配置管理 NAS。采用一款基于 PHP + Postgresql 的开源 Web DMS DocMGR^[6],通过融合 DocMGR 到 NAS 管理系统来改善 NAS 的文档管理功能。DocMGR 可以存储任何类型的文件,能够对最常见类型的文件进行索引,支持文件版本和编辑控制,支持文件工作流,支持细粒度的文件权限控制。

总体融合方案如下:如图 1 所示;通过在 NAS Web 管理端添加配置连接模块和备份恢复模块,实现将 Web DMS 融入 NAS 管理系统。通过 Web 浏览器访问 NAS Web 管理系统,继而进入 Web DMS,通过正确配置连接,即可访问文档管理服务;通过备份恢复模块,实现 Web DMS 数据的备份与恢复,进一步保障数据安全。

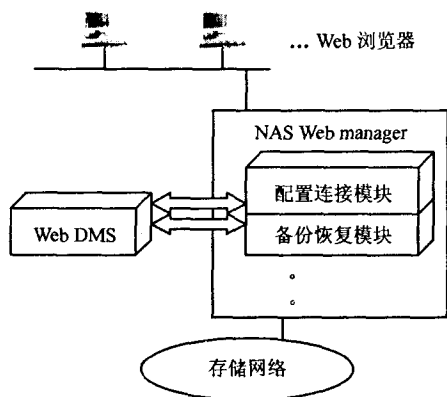


图 1 系统架构示意图

具体实施时,首先考虑到 DocMGR 作为一款开源软件,某些功能在一定环境下与实际使用需求不符,所以先对该 Web DMS 作一些改进。然后对该 Web DMS 编译安装配置包装,再移植到 NAS 管理系统环境下,单独测试运行。最后完善 NAS Web 管理系统,通过添加配置连接模块来实现二者的融合,主要包括完善 Web 页面和后台逻辑处理层,添加处理 Web DMS 的配置选项功能,并负责处理对应数据库以及 Web 服务器的开启和关闭,实现在 NAS Web 管理界面上配置、开启、关闭与连接 Web DMS 功能。由于 DocMGR 本

身没有备份功能,为了防止文档数据遭到意外破坏,保障用户数据安全,所以还要在 NAS 上添加备份恢复模块来保护文档数据。

2 Web DMS 改进

DocMGR 是一款在 Linux 下开发的 Web DMS,需要作一些改进,融入 NAS 后才能更方便使用。经过实际分析,该文档管理系统需要改进以下几个方面:

- (1)多语言支持;
- (2)中文文件缩略图;
- (3)中文全文索引;
- (4)Windows 文件批量上传。

改进后的系统功能模块如图 2 所示,其中虚线部分为添加的功能。

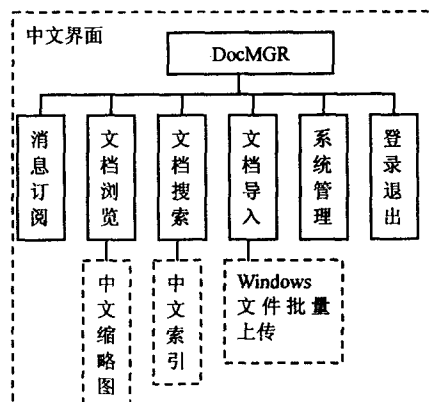


图 2 Web DMS 功能模块图

2.1 多语言改进

DocMGR 仅含有英文界面,如果有个中文界面会更方便使用,因此对其进行汉化处理。Web 应用系统的汉化主要对窗口、标题、菜单、按钮、帮助及警告信息等进行中文化处理。可以考虑两种方案,一种是为每种语言分别准备一套页面文件,这样页面可以直接输出到客户端,占用服务器资源较少。第二种是在需要输出语言文字的地方统一采用语言变量来表示,根据不同的语言赋予不同的值。

DocMGR 用 PHP 开发,采用动态页面方式。DocMGR 有个英文语言包,存放 PHP 语言变量和 javascript 语言变量对应的英文定义。在此基础上,建立对应的中文定义文件,将 PHP 和 javascript 语言变量翻译成合适的中文,例如对于 PHP 语言变量,可以用 define(“语言变量”,“对应中文”),对于 javascript 语言变量,可以用 I18N[“javascript 语言变量名”]=“对应中文”。最后搜索遗漏的语言变量定义,完善中文语言包。

2.2 文件缩略图改进

DocMGR 支持用缩略图和预览图来查看文件,但

是对于中文 text 文档和 word 文档不能很好的支持,不能正确产生缩略图。分析原因是编码问题,改进方法就是要统一编码,使数据传递编码、开发工具编码设置和数据库编码设置做到一致。对于 text 文件,利用 Linux 下的 iconv 工具先将文件编码转换为 UTF8 格式,然后再建立缩略图。对于 word 文档,先利用 anti-word 工具将其转换为 UTF8 编码的 text 临时文档,然后按 text 文档的方式生成缩略图。

2.3 全文索引改进

DocMGR 并不支持中文全文索引,为了方便中文文件索引,需要实现该中文全文索引^[7]功能。研究发现 DocMGR 建立文件索引原理:在对文件进行索引时,利用正则表达式来过滤字符,只保留了字符流中的英文字母和数字,将其他字符全部过滤掉,当然也包括中文字符。由此提出改进方法:在过滤索引内容时,改进正则表达式,添加保留中文字符,这样就可以对中文字符建立索引,达到全文索引中文的目的。改进效果如图 3 所示。

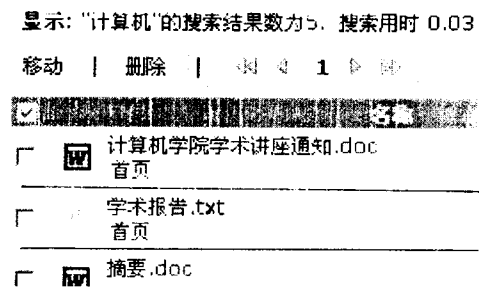


图 3 中文全文索引示例

2.4 批量上传改进

DocMGR 实现了批量上传功能,可以将系统中的文件树导入系统,但是只支持 Linux。因此需要改进批量上传功能,来方便 Windows 用户。改进方法:首先共享需要上传的 Windows 目录,使用 SMB 挂载该文件到 Linux 共享目录下,然后按照处理 Linux 目录树的方式上传该 Windows 目录。实现界面如图 4 所示。

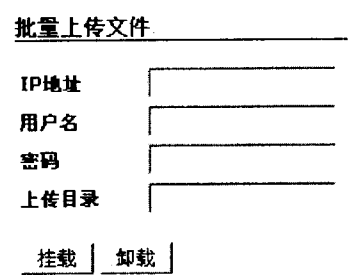


图 4 批量上传 Windows 文件界面

3 Web DMS 与 NAS 管理系统的结合

Web DMS 必须在 NAS 系统环境下能够单独运

行,才可以结合成功。因此,首先要将 DocMGR 移植到基于 Linux 的 NAS 系统环境下测试运行,然后再设计完成 NAS Web 管理端的配置连接模块和备份恢复模块。

3.1 Web DMS 的移植

NAS 管理软件产品一般不会带有编译工具,因此 DocMGR 需要在其他系统环境下编译安装,然后再移植到 NAS。

考虑到 NAS 软件载体的存储空间问题,首先需要对 DocMGR 安装后的文件作一些裁剪。对于 Linux 下的应用程序^[8],安装后一般会生成 bin, lib, share, man, etc 等类似常见目录,裁剪就是要删除不重要的目录,保留运行时必须的 bin, lib 以及配置文件等等。

对于应用程序的移植,可能遇到的问题,一是机器码绑定,有些应用软件安装时绑定了机器码,这样就不能将该软件直接复制到其他机器上运行。二是依赖 lib 库的问题,应用程序正常运行往往需要依赖一些共享 lib 库,用 ldd 脚本可以查看依赖关系。对于不同的系统,所含的 lib 库可能不一样。移植之后要确保找到相同名字相同版本的 lib 库才能保证应用程序正常运行。对于基于 Linux 的 NAS 管理软件,所使用的操作系统往往是经过优化处理的,不可避免缺少一些 lib 文件,在移植 DocMGR 以及相关数据库、Web 服务器等工具软件时,都要检查各个可执行程序是否缺少 lib,如果是,则要同时将缺少的 lib 移植过来。还可能遇到软件运行依赖系统工具的问题。软件在某个系统下运行时用到的基本工具在另一个系统下可能就找不到。在测试运行移植软件时,如果缺少某些系统工具,往往会得到有效的提示信息,此时把缺少的系统工具也一同移植过去,即可使软件正常运行。

3.2 配置连接模块详细设计

配置连接模块设计包括 NAS Web 管理系统、Web 页面设计和后台处理设计。该模块实现了 Web DMS 的配置以及服务的启动停止,并通过链接访问 Web DMS。页面用 html 和 javascript 编写,尽量做到简单易用,如图 5 所示,后台处理由 C++ 实现。实现方案:将 Web DMS 作为一项系统服务,通过在 Web 页面添

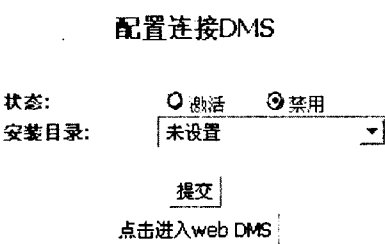


图 5 配置连接界面

加按钮来启动和停止 Web DMS 服务。同时显示 Web DMS 安装目录等信息。如果 Web DMS 没有安装,则通过选择一个共享目录,然后提交,将准备好的 Web DMS 安装包安装到指定目录下。当一切安装配置成功后,通过链接按钮访问到 Web DMS。后台处理程序主要包括两部分:获取页面信息和处理页面请求。

处理代码编写流程如下:

```
DMS::getHtmlContent()//获取页面信息
{ //查看 DMS 服务项是否激活
isServiceEnabled(...)
//获取配置文件中的文件、数据和导出目录路径以及语言设置
GetProfile("file", SYS_CFG_DMS,...)
GetProfile("data", SYS_CFG_DMS,...)
GetProfile("import", SYS_CFG_DMS,...)
GetProfile("lang", SYS_CFG_DMS,...)...
}

DMS::handleSubmit()//处理页面请求
{ //取得页面传来的文件、数据存放路径,导出目录,以及语言
设置,并写入配置文件
iter = pParamMap->find("file")...
iter = pParamMap->find("data")...
iter = pParamMap->find("import")...
iter = pParamMap->find("lang")...
WriteProfile("file", SYS_CFG_DMS,...)
WriteProfile("data", SYS_CFG_DMS,...)
WriteProfile("import", SYS_CFG_DMS,...)
WriteProfile("lang", SYS_CFG_DMS,...)
//执行相关处理命令
fsystem(strCmd...);
//取得页面激活选项
iter = pParamMap->find("status")...
//如果是激活,则启动 DMS,否则停止 DMS
if(iter->second == "yes")
service(SERVICE_START, "DMS"... )
else service(SERVICE_STOP, "DMS"... )
}
```

3.3 备份恢复模块详细设计

实现方案:在页面上列出系统中的共享目录,选中目录显示该目录的相关属性信息。页面上添加备份和恢复按钮,选中一个共享目录,然后点击备份按钮,实现将 Web DMS 的数据备份到该共享目录下;选中一个共享目录,点击恢复按钮,将该目录下已经备份的数据恢复到 Web DMS。处理流程如图 6 所示。在预算较充足时,可以考虑开启 NAS 上的自动备份软件功能,或者购买其他第三方专业备份软件,文中的备份模块给用户多一种备份选择。

4 结束语

NAS 灵活的配置,以及在应用和服务上的不断改进,使得 NAS 在当今存储时代仍扮演重要角色。在 NAS 存储服务广泛部署的背景下,在 NAS 上高效管理文档的需求下,研究了 Web DMS 与 NAS 管理系统结合过程中的一些关键问题。对一款开源 Web DMS DocMGR 进行了多方面的改进,完善了该文档管理软件的编码,对软件编译,安装,移植,配置,最终成功将 DocMGR 与一种基于 Linux 的 NAS 产品结合在一起,方便了 NAS 上的文档管理,已在某档案馆取得了良好的应用效果。

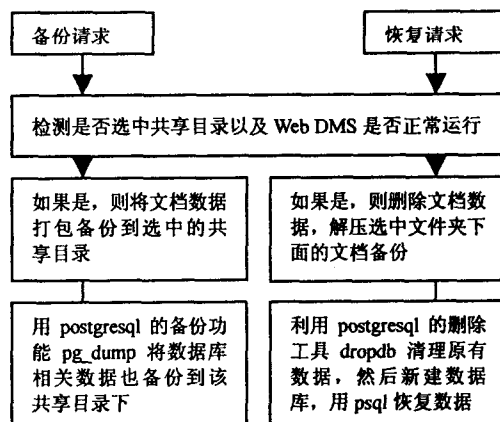


图 6 备份恢复模块处理流程

参考文献:

- [1] Gibson G A. Network attached storage architecture[J]. COMMUNICATIONS OF THE ACM, 2000, 43(11): 37-45.
- [2] 陈 凯, 白英彩. 网络存储技术及发展趋势[J]. 电子学报, 2002, 30(12A): 1928-1932.
- [3] 王彦丽, 陈 明, 陈 峰, 等. 基于 Web Services 企业应用集成的设计与分析[J]. 计算机技术与发展, 2008, 18(9): 212-215.
- [4] 张选平, 谭小鹏, 朱永虎. 面向 Internet 的文档管理系统的设计与实现[J]. 计算机工程与设计, 2004, 25(1): 135-138.
- [5] 刘 丽, 吴秋云, 李 军. 基于 web 的分布式文档管理系统的设计与实现[J]. 计算机工程与科学, 2007, 29(1): 14-16.
- [6] Eric. DocMGR[EB/OL]. 2007. <http://wiki.docmgr.org/index.php/DocMGR>.
- [7] 战 疆, 冯月利, 王 珊. PostgreSQL 中文全文索引技术研究及实现[J]. 华中科技大学学报: 自然科学版, 2005, 33(z1): 213-216.
- [8] 杜 华. Linux 编程技术详解[M]. 北京: 人民邮电出版社, 2007.