

面向科技信息的资源共享服务平台设计

曾 茜¹, 王卓昊², 周明全¹

(1. 北京师范大学 信息科学与技术学院, 北京 100875;

2. 大连理工大学 电子与信息工程学院, 辽宁 大连 116023)

摘要:科技资源的分布式存储与统一管理的要求, 应用系统彼此孤立与信息共享、信息联动之间的矛盾, 这些现实困难与挑战要求一个可用的科技信息资源共享服务平台。提出了一种面向科技信息的资源共享服务平台设计方案, 平台采用混合式的 P2P 体系结构, 由资源汇交系统、资源加工处理系统、资源目录服务系统和日志系统四个子系统组成。资源共享服务平台对科技资源的集成与共享服务发挥了重要作用, 也为其他以科技资源共享为基础的项目提供了软件运行基础。

关键词:资源共享; 元数据; 目录服务; P2P

中图分类号: TP311.1

文献标识码: A

文章编号: 1673-629X(2009)04-0149-05

Design of Scientific and Technological Resource Sharing Platform

ZENG Xi¹, WANG Zhuo-hao², ZHOU Ming-quan¹

(1. College of Information Science and Technology, Beijing Normal University, Beijing 100875, China;

2. School of Electronic and Information Engineering, Dalian University of Technology, Dalian 116023, China)

Abstract: Proposes the requirement to build scientific and technological resource sharing platform. Describes that resource sharing platform consists of four subsystems, which are resource gathering subsystem, resource processing subsystem, resource directory service subsystem and log subsystem. Further, explicates clearly the functions and framework of the four subsystems. Also shows that the resource sharing platform adopts a hybrid P2P architecture, so as to achieve the distributed storage and unified management of scientific and technological resources.

Key words: resource sharing; metadata; directory service; peer to peer

0 引言

2004 年, 我国启动了“国家科技基础条件平台建设”项目, 其目的是要充分运用信息、网络等现代技术, 对科技基础条件资源进行战略重组和系统优化, 以促进全社会科技资源高效配置和综合利用, 提高科技创新能力^[1]。而我国科技信息资源的有效整合、共享和服务是其重要建设内容之一。

我国有 34 个省市自治区, 各地区都有相应的科技信息工作部门, 传统的科技资源共享交换是依靠组织体系或通过邮件、移动存储设备、CD、DVD 等媒介来完成的, 这样的资源共享范围有限, 交换过程耗时耗力, 并且传递的资源内容分散, 难以实现统一管理与可靠存储。随着科技信息服务建设工作的深入开展, 各

地区都逐渐建立了自己的科技资源数据库和应用系统, 通过对我国各省市信息科技资源数据库和相关应用系统的调研发现, 由于缺乏统一的规划和管理, 各地区科技资源分类体系不一, 各个资源数据库和应用系统彼此孤立, 相互之间难以实现资源共享和信息传递, 致使各个应用系统只能是孤岛式的运行, 无法实现资源间、系统间的资源共享和信息联动。

这些现实困难与挑战要求一个可用的科技信息资源共享服务平台, 它可以提供:

(1) 便捷的资源汇交工具, 支持异构科技资源数据库的汇交, 以使资源提供者可以轻松地提交他们想要共享的科技资源;

(2) 汇交资源的加工处理, 以保证共享资源的科学性、有效性;

(3) 资源的恰当分类与存储, 为共享资源提供简单快捷的访问方式;

(4) 合理的系统体系结构, 支持资源的分布式存储与统一管理。

收稿日期: 2008-08-26

基金项目: 国家科技基础条件平台项目(2005DKA63904)

作者简介: 曾 茜(1986-), 女, 江西人, 硕士研究生, 研究方向为信息技术与应用; 周明全, 教授, 博导, 研究方向为计算机可视化技术、网络信息处理技术等。

在实施“国家科技基础条件平台应用服务支撑系统”子项目的过程中,设计并实现了面向科技信息的资源共享服务平台,为科技资源的集成共享提供软件平台支撑。

1 资源共享服务平台系统设计

1.1 总体结构

资源共享服务平台由资源汇交系统、资源加工处理系统、资源目录服务系统和日志系统四个子系统构成。根据项目对用户统一管理的要求,资源共享服务平台的用户管理采用项目统一的用户注册与身份认证系统以及授权与访问控制系统,根据系统提供的访问接口进行交互,实现基于角色的用户授权与访问控制,在此则不作具体描述。资源共享服务平台的功能结构如图 1 所示。

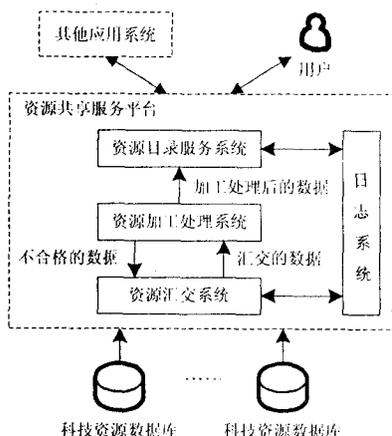


图 1 资源共享服务平台功能结构图

资源汇交系统是实现资源共享的关键,它支持资源提交者对科技资源进行定制,配置科技资源元数据信息,提供科技资源数据库的访问接口;支持资源提交者进行目录融合操作,以实现科技资源与系统科技资源标准分类目录的关系映射,达到资源分类体系的统一。资源加工处理系统支持资源管理者对汇交的科技资源数据进行标引和审核,从而保证共享资源的科学性、有效性。资源目录服务系统支持系统管理者对系统科技资源标准分类目录进行增加、删除、修改等管理工作,简化目录管理工作;提供科技资源目录访问接口,支持用户按目录层级浏览汇交的有效科技资源。日志系统记录科技资源汇交的相关信息,如提交用户、数据容量、提交日期、用户 IP 等;另外还记录用户的访问或增加、删除、修改等操作,为今后用户个性化服务以及相关统计分析工作提供数据支持。

1.2 资源汇交系统

资源汇交系统是资源加工处理系统和资源目录服务系统的前提,资源提交者通过资源汇交系统进行相

关配置,实现其科技资源数据库的共享接入与目录融合。资源汇交系统主要包括资源定制、资源编辑、目录融合、资源封装与传输、资源解析、资源更新以及操作日志等模块,如图 2 所示。

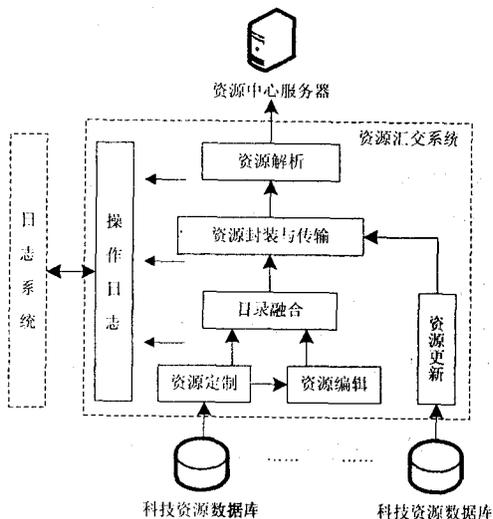


图 2 资源汇交系统功能模块图

资源定制模块用于建立与科技资源数据库的连接,并完成元数据配置。其流程为:资源提交者选取本地要接入共享的科技资源数据库,根据数据库类型配置相应的数据库驱动,并填写连接数据库所必需的参数;测试与本地科技资源数据库的连接,测试成功则进行元数据信息的填写,否则返回前一步即数据库连接配置^[2];选择资源汇交模式(集中模式和分布模式),集中模式则将科技资源数据库的全部数据连同所提取的数据库信息以及元数据信息一起提交到资源中心,而分布模式则只提交所提取的数据库信息以及元数据信息。此模块提取的数据库信息、元数据信息以及所选择的汇交模式信息均存储在 XML 文件中^[3],同时将资源提交者的用户 ID、数据库大小、日期、IP 地址等信息写入操作日志。

资源编辑模块用于资源提交者对已定制的资源进行修改,可更改数据库连接配置以及元数据信息。更改完成后对原 XML 文件进行更新,同时将用户 ID 以及所作的更改操作写入操作日志。

在之前获取的数据库信息及元数据信息的基础上,目录融合模块支持资源提交者建立其科技资源目录,然后实现与系统科技资源标准分类目录的关系映射,从而将资源提交者汇交的科技资源数据映射到统一的分类体系中,实现科技资源的统一分类。目录融合完成后,科技资源数据表信息及其所属目录信息同样以 XML 文件存储。

针对集中汇交模式,资源更新模块用于实现资源中心与本地科技资源数据库的同步,将本地科技资源

数据库中的原始数据的修改以及新增数据更新到资源中心,保证资源中心与本地科技资源数据的一致性。新增的数据所属目录与其所属数据表保持一致。

资源封装与传输模块用于对包含各种配置信息的XML文件,如数据库信息、元数据信息、汇交模式信息和目录融合信息等,及科技资源数据(针对集中汇交模式)进行格式转换和统一封装,建立与资源中心服务器的通信连接,将封装好的数据传至资源中心服务器。

数据资源解析模块用于在资源中心服务器端对所接收到的数据包进行解析,获取由数据包封装的相关配置文件,包含数据库信息、元数据信息、汇交模式信息和目录融合信息等,以及数据文件(针对集中汇交模式),通过配置文件解析、标准格式映射、实体数据存储等过程,最终实现科技资源的汇交。

操作日志模块为资源汇交系统与日志系统的接口。资源汇交系统中用户(资源提交者)的相关信息、所提交资源的相关信息以及操作信息通过操作日志模块记录,并按照一定格式存储,提供给日志系统。

1.3 资源加工处理系统

资源加工处理系统是资源共享服务平台的重要组成部分。该系统负责对资源汇交系统所汇交的数据进行加工处理,基于 workflow 技术实现资源加工处理的流程化管理,为资源目录服务系统提供有效、科学、权威的科技资源数据。资源加工处理系统主要包括资源审核模块和数据标引模块,如图3所示。

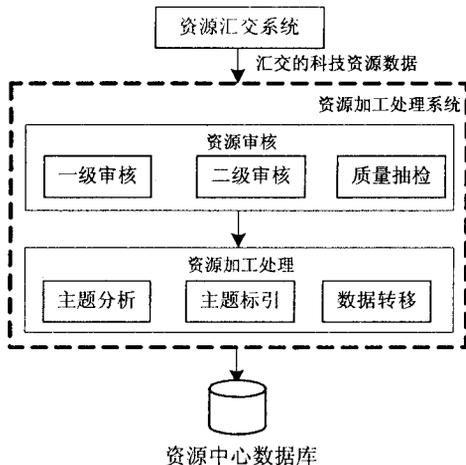


图3 资源加工处理系统功能模块图

资源审核模块用于对汇交到资源中心服务器的信息和数据进行审核,实现汇交资源的质量控制。资源审核模块汇总待审核的任务,根据审核级别的不同呈现给审核员不同的审核任务列表,最终使得各级审核都通过的资源能够被用户访问。

资源加工处理模块用于对通过审核的科技资源数据进行主题分析和主题标引,支持标引的增加、删除、

修改和浏览;此外,此模块还负责数据的定时转移,在设定的时间段将已完成主题标引的资源数据转移到相应的存储位置或存储数据库,以供用户访问。

1.4 资源目录服务系统

资源目录服务系统是实现资源共享服务的核心。在相关科技资源分类体系标准规范的指导下,该系统负责建立科技资源标准分类目录,并支持对分类目录的增加、删除和修改等管理行为;将汇交的有效科技资源数据以目录导航的方式提供给用户访问,同时提供目录服务访问接口,为其他应用系统集成资源共享服务平台的建设成果提供支持。资源目录服务系统包括目录建立、目录管理、访问接口、目录导航以及操作/访问日志等模块,如图4所示。

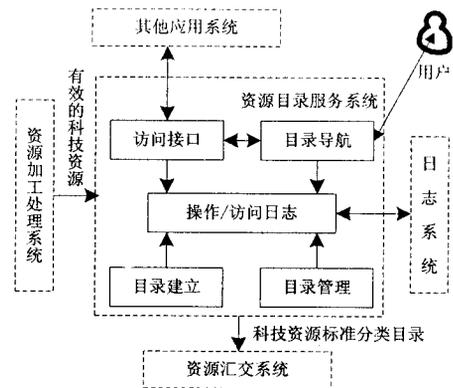


图4 资源目录服务系统功能模块图

目录建立模块用于建立系统科技资源标准分类目录,而采用的分类体系标准均按照项目标准规范的内容要求执行,即为同时采用国民经济行业分类、社会经济目标分类、学科分类、地域分类和国家科技基础条件平台分类五大分类体系。目录分类编码均按照国家标准执行。

目录管理模块用于对建立的科技资源标准分类目录进行增加、删除和修改等操作,使系统具有一定的扩展性。资源目录服务系统所建立的科技资源标准分类目录将提供给资源汇交系统,用于目录融合模块中。

访问接口提供对汇交的有效科技资源标准分类目录的访问接口,为其他应用系统集成资源共享服务平台的建设成果提供通道。目录导航模块响应用户的请求,为用户提供统一的科技资源目录视图,通过与访问接口的交互,使用户可以按目录层级浏览资源,从而实现资源的分类导航。

操作/访问日志模块记录管理员用户的具体目录管理操作,如删除了哪个分类体系下的哪个目录节点,使管理操作都有据可查;同时,还记录用户对资源的访问情况,为今后用户的个性化服务以及相关统计分析工作提供数据支持。

1.5 日志系统

日志系统是为了满足资源共享服务平台的系统管理需求,它为系统管理员提供基本的操作历史数据和访问历史数据,并按照不同的类别将日志数据以友好的方式呈现给系统管理员。此外,它还支持基于用户、日期等关键字的日志数据查询,使系统管理员能更加快捷地查找所需记录。日志系统主要包括资源汇交日志、目录操作日志、资源访问日志以及日志检索等模块,如图 5 所示。

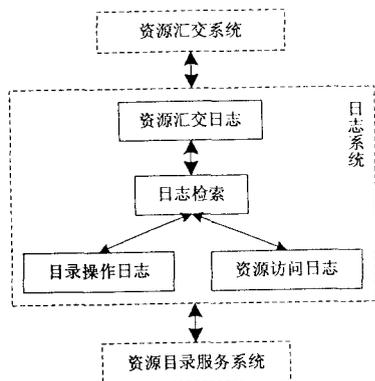


图 5 日志系统功能模块图

资源汇交日志模块用于呈现资源汇交的日志信息,主要包括资源提交者的用户 ID、数据库大小、提交日期、IP 地址等信息。目录操作日志模块用于呈现目录操作的日志信息,主要包括管理员的用户 ID、IP 地址、操作时间、操作类型等信息。资源访问日志用于呈现资源访问情况的日志信息,主要包括用户 ID、IP 地址、访问时间、访问的资源等信息。而日志检索模块则用于快速查找所需的日志信息,可以根据用户名查找其关联的所有日志信息,或查找某一时间段内的所有日志信息。

2 资源共享服务平台体系结构

2.1 P2P 简介

根据文献[4],P2P 可定义为:

定义 1 P2P 即对等计算,是一种新的分布式计算模式。这种计算模式组成的网络中,节点之间的地位是对等的,网络中的所有节点共享它们的部分硬件或软件资源;这些共享资源能被其他对等节点直接访问而无需经过任何中间节点;每个节点既可以是数据和服务的提供者,也可以是数据和服务的需求者。

P2P 系统在实际的产品实现中存在着多种不同的结构,如:Freenet^[5]、Gnutella^[6]等采用的所有节点完全对等的纯 P2P 结构,Napster^[7]采用的带有中央服务器的混合 P2P 结构以及 Kazaa^[8]采用的具有多个超级节点的 P2P 结构等。

根据文献[9],混合 P2P 结构可定义为:

定义 2 如果一个分布式网络结构满足下面的两个条件,则可以称为混合 P2P 结构:①满足定义 1;②必须有一个中央节点(目录节点)专门提供服务,此节点仅提供服务而不向其他节点提出请求。

从本质上说这种结构并不是纯粹的 P2P,因为这里存在一个中央节点,但是这里的中央节点同于传统 C/S 模式中的服务器,它仅仅负责维护整个系统中所有内容的目录索引,并不存放真正的文件内容。当节点加入系统时,需要向中央节点进行注册登记节点的有关信息和所共享文件的目录。当需要查询某个文件时,对等节点会向中央节点发出文件查询请求。中央节点进行相应的检索和查询后,会返回符合查询要求的对等节点地址信息列表。查询发起的对等节点接收到应答后,会根据网络流量和延迟等信息进行选择,与合适的对等节点建立连接,并开始文件传输。

2.2 基于混合 P2P 的资源共享服务平台体系结构

资源共享服务平台将部署在各个资源节点上,最终整个平台将由一组资源节点服务器和一个中央服务器组成。一个资源节点服务器通常为某一个拥有资源的科研机构或部门所有,由他们负责管理运行,这样可以更好地对资源数据进行收集、更新、存储和管理。而通过资源共享服务平台提取其元数据信息,在中央服务器存储并建立索引,这样用户就可以按目录浏览资源节点服务器上的资源。

与 MP3 共享软件 Napster^[7]、科学实验数据共享软件 SciPort^[10,11]相似,资源共享服务平台采用混合式的 P2P 体系结构来共享和集成科技资源数据,如图 6 所示,所有资源节点服务器就是通过混合式的 P2P 体系结构连接在一起。这种体系结构用一台中央服务器来存储各节点资源的元数据信息,而各个资源节点服务器则负责存储科技资源数据。这样的混合式 P2P 体系结构可以更好地管理元数据,也能够支持进一步的资源查询检索。不过,在资源共享服务平台的体系结构中,中央服务器还可以存储一部分科技资源数据。

每一个资源节点服务器都安装着一套完全相同的软件系统即资源共享服务平台,资源节点服务器可以作为服务器独立运行。科技资源数据存储在本地,而每当增加元数据信息时,元数据首先添加到本地服务器(即资源节点服务器),然后再发送到中央服务器。而目录信息的增加和修改也同样如此。

中央服务器首先是一台目录服务器。与传统的文件共享系统(如 Napster^[7])不同,它除了保存着文件名及其存储位置信息外,还包含有更丰富的信息。中央服务器存储着各个资源节点服务器发来的元数据,并

根据资源节点服务器发来的目录信息,建立起一个覆盖所有节点资源数据的统一目录视图,在用户权限认证通过的情况下,用户可以连接到资源节点服务器,获取具体的科技资源数据。

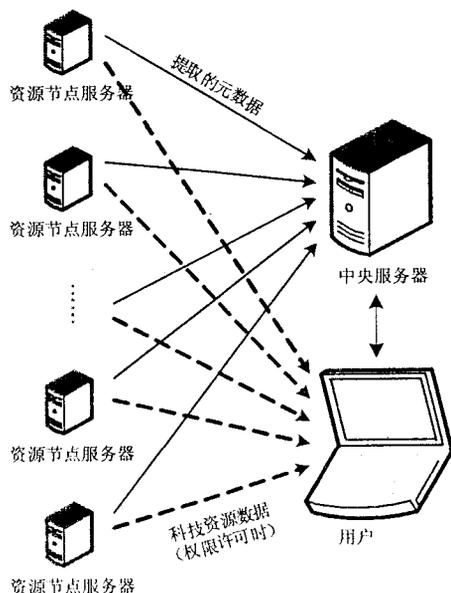


图 6 混合式 P2P 体系结构

3 结束语

结合我国科技资源的建设现状,给出了一种面向科技信息的资源共享服务平台设计方案。系统采用混合式的 P2P 体系结构,满足了科技资源的分布式存储与统一管理的需求,而功能上也覆盖了资源的汇交、加工处理和目录服务等实现资源共享所需的过程。正是在项目实施过程中设计并实现了这一资源共享服务平台,对科技资源的集成与共享服务发挥了重要作用,同时也为其他以科技资源共享为基础的项目提供了软件运行基础。

参考文献:

[1] 科技部,发展改革委,教育部,财政部. 2004-2010 年国家

(上接第 148 页)

参考文献:

[1] Yaghmour K. 构建嵌入式 Linux 系统[M]. 韩存兵,龚波 改编. 北京:中国电力出版社,2004.
 [2] Aleph One Ltd. Embedded Debian, Yaff s : A NAND2Flash Filesystem[OL]. 2002. <http://www.aleph1.co.uk/yaffs/>.
 [3] 熊伟,董金明. 嵌入式 Linux 中根文件系统的实现[J]. 电子测量技术,2007(7):25-30.
 [4] 马忠梅,李善平,康慨,等. ARM & Linux 嵌入式系统教程[M]. 北京:北京航空航天大学出版社,2004.

科技基础条件平台建设纲要[EB/OL]. 2004-09-15. [2008-08-01]. <http://www.most.gov.cn/gjkjtcjptjs/zcfg/wj/200409/t20040915-15767.htm>.

[2] 孙晓芬,刘淮松,董晶. 虚拟数据视图的设计与实现[J]. 计算机工程与设计,2006,22(27):4337-4339.
 [3] 何敬刚,霍宏,方涛. 基于 MVC 模式空间元数据目录服务的实现[J]. 计算机工程与应用,2006(13):165-167.
 [4] Schollmeier R. A definition of peer-to-peer networking for the classification of peer-to-peer architectures and applications[C]//In: Proceedings of the First International Conference on Peer-to-Peer Computing. Linkoping, Sweden: IEEE Computer Society,2001.
 [5] Clarke I, Sandberg O, Wiley B, et al. Freenet: A Distributed Anonymous Information Storage and Retrieval System[C]// In Proceedings of the ICSI Workshop on Design Issues in Anonymity and Unobservability. Berkeley, California: [s. n.], 2000.
 [6] Ripeanu M. Peer-to-peer architecture case study: Gnutella network[R]. Technical report. USA: University of Chicago, 2001.
 [7] Napster Inc. The napster homepage [EB/OL]. 2001. In <http://www.napster.com/>.
 [8] Good N S, Krekelberg A. Usability and privacy: a study of Kazaa P2P file-sharing[C]//In Proceedings of the CHI 2003 conference on human factors in computing systems. Fort Lauderdale, Florida: [s. n.], 2003.
 [9] 王珊,张新宇. Peer-to-Peer 数据共享研究[J]. 计算机应用与软件,2003,20(11):1-4.
 [10] Wang Fusheng, Liu Peiya, Pearson J. et al. Experiment Management with Metadata-based Integration for Collaborative Scientific Research[C]//In Proceedings of the 22nd International Conference on Data Engineering. Atlanta, Georgia: [s. n.], 2006.
 [11] Wang Fusheng, Bourgué P E, Hackenberg G. SciPort: an adaptable scientific data integration platform for collaborative scientific research[C]//In Proceedings of the 33rd international conference on Very Large Data Bases. Vienna, Austria: [s. n.], 2007.
 [5] Corbet J, Rubini A, Kroah-Hartman G. LINUX 设备驱动程序[M]. 魏永明,耿岳,钟书毅,译. 北京:中国电力出版社,2006.
 [6] 张方樱. 构建嵌入式 Linux 的根文件系统[J]. 实验室科学, 2007(6):28-30.
 [7] 卢剑翔,刘成安,胡和智,等. 基于 S3C2410 的 Cramfs 根文件系统的移植[J]. 微机计算机信息,2006,22(11):22-25.
 [8] 杨延军. 用 Busybox 制作嵌入式 Linux 的文件系统[J]. 单片机与嵌入式系统应用,2005,1:17-19.
 [9] 张勇. 嵌入式 Linux 下 JFFS2 文件系统的实现[J]. 计算机技术与发展,2006,16(4):175-180.