

科技信息资源共享服务平台设计初探

陈 隽

(江苏淮安市行政学院,江苏淮安 223001)

摘要:科技信息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4个子系统构成。该平台体系的建设采用混合式的P2P体系结构。这种体系结构是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。

关键词:资源共享;元数据;目录服务;P2P;信息共享;科技信息资源共享平台

中图分类号: G311 **文献标识码:** A **DOI:** 10.3772/j.issn.1674-1544.2009.06.003

1 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。

我国各省市自治区和特区,都设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各自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构建的科技信息资源共享服务平台,能够提供以下服务: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共享的科技资源;汇交资源的加工处理,以保证共享资源的科学性、有效性;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;合理的系统体系结构,支持资源的分布式存储与统一管理。

2 资源共享服务平台系统设计

2.1 总体结构

科技信息资源共享服务平台应由资源汇交系统、资源加工处理系统、资源目录服务系统和日志系统4个子系统构成。根据项目对用户统一管理的要求,科技信息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

作者简介:陈隽(1975-),女,计算机学士,讲师,研究方向是政府信息化、电子商务。

收稿日期:2009年8月21日。

行交互,实现基于角色的用户授权与访问控制。资源共享服务平台的功能结构如图 1 所示。

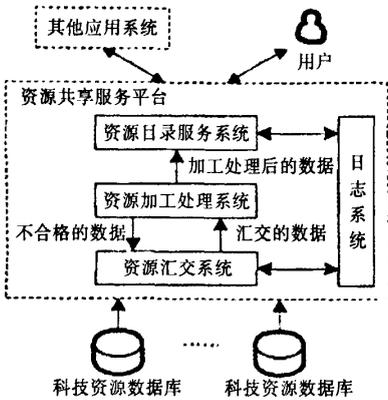


图 1 资源共享服务平台功能结构图

2.2 资源汇交系统

资源汇交系统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。

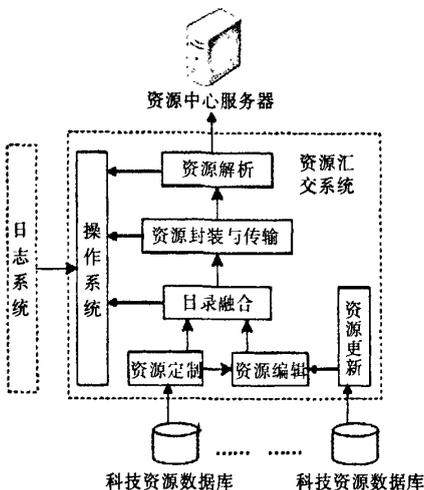


图 2 资源汇交系统功能模块图

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 以及所做的更改操作写入操作日志。

在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)以及科技资源数据(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。

操作日志模块是资源汇交系统与日志系统的接口。在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。

2.3 资源加工处理系统

资源加工处理系统负责对资源汇交系统所汇交的数据进行审核和加工处理,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(图3)。

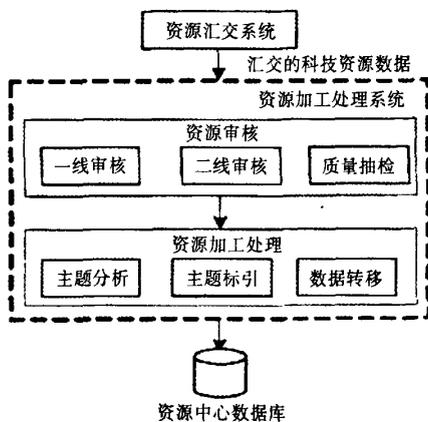


图3 资源加工处理系统功能模块图

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。对于资源审核模块汇总待审核的任务,根据审核级别的不同,审核员将接到不同的审核任务列表,最终通过各级审核都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。

2.4 资源目录服务系统

在相关科技资源分类体系标准规范的指导下,资源目录服务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(图4)。

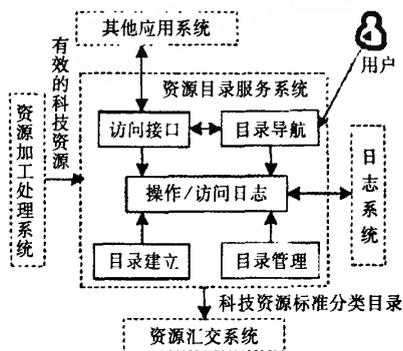


图4 资源目录服务系统功能模块图

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。

访问接口模块是为汇交的有效科技资源标准分类目录提供的,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。

2.5 日志系统

日志系统是为了满足资源共享服务平台系统管理的需求,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,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(图5)。

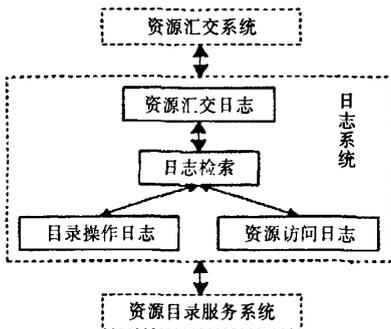


图5 日志系统功能模块图

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。

3 资源共享服务平台体系结构

计算机对等网(Peer-to-peer network, P2P)技术能够对元数据信息进行管理,并进行资源查询检索,是目前流行于国际计算机网络技术研究领域的一个热点。在设计资源共享服务平台时候,应充分利用这种先进的技术。

3.1 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,也可以是数据和需求者。

从P2P的定义可以看出,网络中的节点既可以获取其他节点的资源或服务,同时又是资源或服务的提供者,即兼具客户机和服务器双重身份,因此P2P是一种分散的、分布式的资源管理模型。

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,Kazaa^[8]采用的具有多个超级节点的P2P结构等。

3.2 混合P2P

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)除去一个节点外,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不同于传统C/S模式中的服务器,其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。

3.3 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(图 6)。

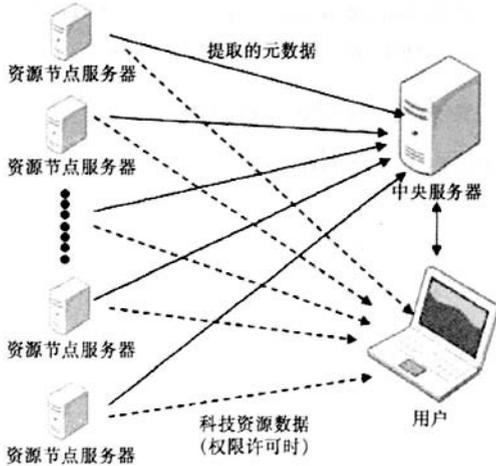


图 6 混合式 P2P 体系结构

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 Napsler)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。

4 结 语

根据我国科技资源的建设现状,科技资源的分布式存储与统一管理要求的矛盾,以及应用系统彼此孤立与信息共享、信息联动之间的矛盾,本文对科技信息的资源共享服务平台设计进行了探索,提出了一种面向科技信息的资源共享服务平台设计模式。该方案采用混合式的 P2P 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。在实际操作过程中也取得了比较理想的应用效果。希望这种体系结构设计能为其他以科技资源共享为基础的项目提供软件运行基础。

参考文献

- [1] 科技部,发展改革委,教育部,财政部. 2004-2010 年国家科技基础条件平台建设纲要 [EB/OL]. [2008-08-01]. <http://www.most.gov.cn/gjkjje-tjptjs/zcfc/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,崔宏,力涛. 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] Schollmicer R. A Definition of Peer-to-peer Network for the Classification of Peer-to-peer Architectures and Applications[C] // Proceedings of the First International Conference on Peer-to-Peer Computing. Linkoping, Sweden: IEEE Computer Society, 2001.
- [5] Clark I, Sandberg O, Wiley B, et al. Freenet: A Distributed Anonymous Information Storage and Retrieval System[C] // 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]. [2007-01-08]. <http://www.napster.com/>.
- [8] Good N S, Krekedberg A. Usability and Privacy: a Study of Kazaa P2P File-sharing[C]//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]//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]//Proceedings of the 33rd International Conference on Very Large Data Bases. Vienna, Austria: [s. n.], 2007.

Design of Science and Technology Information Resources Shared Services Platform

Chen Jun

(Huai'an Administrative College, Huai'an 223001)

Abstract: System resource sharing service platform is constituted by four subsystems, i. e. resources remitted by the system, resource processing systems, resources, directory services system and log system. This paper discusses the program which uses the P2P hybrid architecture, this architecture with a central server stores the resources of each node meta-data information, while each resource node server is responsible for storing data on scientific and technological resources. This hybrid P2P architecture can be better management of metadata, can also support further resource query retrieval.

Keywords: resource sharing, metadata, directory services, P2P, information sharing, S&T information resources sharing platform

第五届中国科技政策与管理学术年会在西安举行

本刊讯 2009年10月16-17日,以“经济危机背景下的创新政策与管理”为主题的第五届中国科技政策与管理学术年会暨研究会理事会在西安交通大学举行。来自清华大学、北京大学、浙江大学、武汉大学、西安交通大学、大连理工大学及中国科学院、中国科学技术信息研究所、上海市科学学研究所等众多高校及科研院所的200余名科技政策与管理领域的研究人员参加了本届年会并进行了交流。本次年会共设科学学基本理论与学科建设、经济变化与科技创新、国家创新体系建设与科技体制改革、自主创新与企业创新能力、区域创新与社会发展等5个分会场,共有40余篇论文进行了宣讲与交流,许多思想与观点引起了参会人员的共鸣。来自东北师范大学政法

学院的许志峰教授提出了大科学观雷达模型,设计了以数学科学、自然科学、生命科学、社会科学、思维科学、哲学科学6大模块为基础的中国大学通识教育课程体系;来自西安交通大学管理学院的胡平研究团队就我国国家杰出青年科学基金资助效果的评价方法、评价指标和主要结论等进行了阐述;来自西交利物浦大学管理系曹瑄玮以鲁尔区钢铁产业集群为例对组织学习与集群发展路径演化进行了分析,发现区域内核心组织的组织学习对于整体区域的发展路径具有显著的影响作用等。其他比较受关注的议题还有科学计量、知识发现、技术创新、产业创新等。大会研讨内容既涉及理论探讨,又不乏实践指导,达到了为构建创新型国家提供理论参考和决策借鉴的目的。