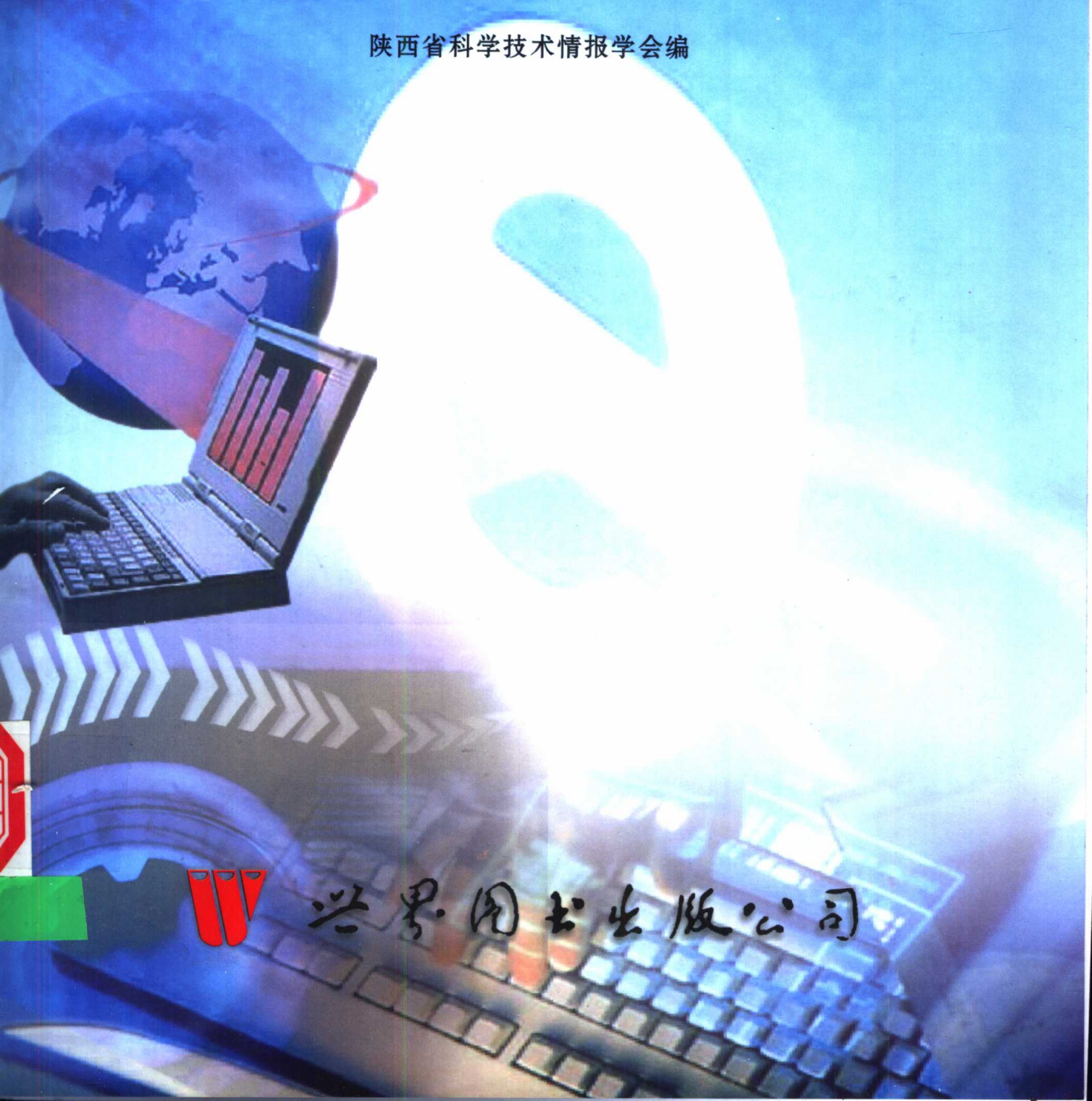


信息 · 网络 · 咨询

INFORMATION INTERNET REFERENCE

陕西省科学技术情报学会编



世界图书出版公司

信息·网络·咨询

主 编

叶春峰 余诗武 杨顺云



世界图书出版公司

西安·北京·广州·上海

信息·网络·咨询

主 编 叶春峰 余诗武 杨顺云
责任编辑 侯清高

出版发行 世界图书出版西安公司
地 址 西安市南大街17号 邮编 710001
电 话 029-7279676 7233647 (发行部)
传 真 029-7279675
E - m a i l wmcxian@public.xa.sn.cn
经 销 各地新华书店
印 刷 陕西省乾兴印刷厂
开 本 787×1092 1/16
印 张 22.125
字 数 400千字

版 次 2002年12月第1版 2002年12月第1次印刷
书 号 ISBN 7—5062-2642-1/Z.104
定 价 36元

前 言

人类迈向 21 世纪之后,出现了若干引人注目的新概念和发展趋势,如信息高速公路、知识经济、网络文化、虚拟图书馆、数字时代等。现代科技被不断地应用到社会各行各业,在如此高速发展的信息时代,图书情报事业应用了网络、数字等先进技术之后,也取得了可喜的成绩。为了全面迅速地提高图书情报及信息工作在新形势下的服务能力和影响力,陕西省科技情报学会于 2002 年 11 月 15 日至 17 日特别筹划召开了陕西省情报学、图书馆学学术论文研讨会。

本次会议论文来自我省 7 个城市的情报机构、公共图书馆、高等院校、科研院所情报室和北京 2 所高等院校图书馆。论文内容涉及到图书情报专业各个方面,如:建立情报文献数据库、数字图书馆的建设、信息检索、网络资源建设和优化文献服务等。

会议先后共收到论文 210 余篇,经著名情报学专家蔡董老先生等人认真审阅,特别精选出 100 篇论文并加以大力斧正录用,其中优秀论文 38 篇,会议交流论文 62 篇。这些论文从不同程度反映了我国图书情报事业发展的现状,极具代表性。有些论文视角新颖、见解独特,对当前图书馆情报工作存在的热点、难点问题进行了探讨,并且提出了富有创新意义的新观点、新思路和新举措,具有很大的参考价值。

从这次学术活动所收集的 210 余篇论文和研讨会上各地代表的积极交流,我们充分认识到,数字化信息革命作为一场新的技术革命,在逐渐改变人们的工作行为方式和价值观念。我省图书情报机构正在发生巨大变化,并面临着这场革命的巨大挑战。无论是体制改革、信息资源开发、数字化、网络化建设,还是情报理论与基础研究等方面都出现了新的面貌。

本书主要由信息与网络技术、信息咨询与服务、机构管理改革和资源建设、信息队伍建设、信息学研究五个部分组成,分别由李涪、翟美珠、张幸芝、张小曼和蔡勤等编辑校对。本书各部分主题突出、观点新颖、条理清晰,经过多次编辑校对,终于正式出版发行。在此,向所有征文作者表示衷心的感谢!向为论文评审和编辑工作付出辛劳的评委们和编辑们表示诚挚的谢意!

叶春峰

目 录

◇ 信息与网络技术 ◇

数字图书馆业务系统设计实践	王建平(1)
磁盘存储技术在电子阅览室中的应用	郭 梅(4)
网络数据库的质量评价	张幸芝 宓 欣(9)
数据仓库为图书馆服务模式转变提供技术支持	王晓丽 赵宇欣(12)
浅评身份识别在联合实时咨询中的实现	陈楠楠(15)
浅评联合虚拟咨询中的咨询信息管理系统	陈楠楠 韩 萌(18)
利用 ODBC 实现 Notes 系统与外部 SQL Server 数据库通信	李 娟 杨 峰(21)
现代环境下的图书馆网络安全初探	阎晓弟(25)
论图书馆自动化系统建设的标准化	郭小民(29)
智能门禁系统在高校图书馆中的应用	肖 华 朱 平 韩 华(32)
国内图书馆自动化集成管理系统分析与展望	高建中(34)
数字图书馆建设:认识与实践	邵小玲(39)
论数字图书馆与高校图书馆	赵玉光 刘雪梅 贺小平(42)
数字图书馆浅议	张 林 刘 霞(46)
高校图书馆网站建设小议	江 汶(50)

◇ 信息咨询与服务 ◇

浅谈专利查新的标准化、规范化及新颖性问题	孟小衍(53)
信息服务创新浅探	叶春峰(56)
手工与网络词典的利用	孙金立 王 栋 李路路(61)
高校图书馆信息服务工作发展与改革	张小曼(65)
浅论电子阅览室服务	杨 芳 薛秦芬(69)
信息时代对加强高校图书馆服务功能的思考	刘秀兰(72)
系列性研究课题查新方法探讨	龚庆侠(74)
医药卫生科技查新常用六种光盘数据库特点比较与评析	龚庆侠(80)
信息咨询业的现状及发展趋势	仇洪江(86)
21 世纪图书馆的信息服务模式	李桂影 葛郁葱(89)
浅析陕西省图书馆读者的阅读行为	孙红侠(92)
浅述新形势下高校图书馆阅览服务工作	李亚莲 王春利(95)
充分利用 NLM 的信息资源	李路路 金 科 王 栋(98)
关于生物医学信息手工检索的教材内容改革	孙金立 马炳厚 李路路(101)
网络环境下图书馆的主动服务	雒 虹 艾润莲(104)
网络环境下图书馆的信息服务	陆 地(107)

ISI Web of Knowledge 平台上新的网络数据库·····	孟小衍(110)
院系资料室向学生开放的尝试·····	常 勇(113)
充分利用阅览室加强高校图书馆文献服务·····	程雪艳(116)
网络环境下图书馆参考咨询的思路与对策·····	来雪玲(119)
大学图书馆参考咨询工作思路·····	王 嫵(122)
论信息用户的需求规律·····	白筱芬 王海萍 李志强(126)
浅析随书光盘的使用与管理·····	赵 辉(130)
新时期图书馆读者服务工作探讨·····	郑怀远(132)
浅探现代环境下图书馆服务新格局·····	苏春萍 杨 璐(134)
读者服务应从小处着眼·····	高云燕(137)

◇机构管理改革和资源建设◇

试论自由主题词·····	李 红(140)
《中国图书馆分类法》第四版 U 类修正之我见·····	刘改玲 王 涛(143)
世界汉语教学学科领域研究的人才培养·····	周志玉(146)
谈谈 CIP 存在的问题及对策·····	冯永财 惠海龙(149)
网络环境下的图书馆计算机编目·····	葛郁慈 李桂影(154)
试论网络信息的组织·····	王 琦(156)
网络环境下大学图书馆服务功能和服务模式的变革研究·····	李桂芳 徐家燕(160)
高校图书馆的知识化管理与知识的管理·····	刘崇学(165)
论图书馆信息系统的构成·····	仇洪江(168)
新时期教育研究型大学图书馆文献信息资源建设思考·····	王公尚(171)
网络环境下外文期刊资源建设与服务·····	梁 云 罗广玲(174)
西部大开发与西部高校图书馆建设·····	李小鸽(176)
西部大开发和图书馆的软管理·····	刘 静 张晓刚(180)
基于 Lotus Domino/Notes 的图书馆人事信息管理系统的实现·····	杨 峰(184)
论新时期大学图书馆馆长的职能·····	闫 鸿(190)

◇信息队伍建设◇

图书馆员继续教育的失误与思考·····	刘卫武 刘 亮(193)
新时期大学生思想状况研究及对策·····	叶会显(196)
浅议学校图书馆员的人格效应·····	袁素瑛(201)
由数字图书馆谈图书馆员队伍建设·····	尹 屹(204)
论信息技术教育对图书馆信息资源建设的促进作用·····	徐家燕 李桂芳(208)
图书馆流通人员语言艺术初探·····	詹 东 王 岚(213)
浅议数字化图书馆对馆员的素质要求·····	王春利 蔡 勤 李亚莲(215)
论网络环境下图书馆员的继续教育·····	翟美珠(218)
论新世纪对高校图书馆馆员的素质要求·····	李 瑾 冯瑞娟(221)
图书信息人才培养内容及发展趋势·····	刘志云 牛红亮(223)

数字图书馆的发展特征及馆员素质的提高	石 蕾 王丽珍(227)
网络环境下医学图书馆员的素质	张 林 石 蕾 翟中会(229)
图书馆应为培养创新人才营造良好的环境	蔚今华(233)
论知识经济时代下图书馆员的素质	杨 薇(237)
高校图书馆期刊核心馆员的角色定位	赵 玲(240)
图书馆工作人员的继续教育和培训	宁亚玲(243)
试论高校图书馆馆员的继续教育	戴晓芹(247)
网络环境下图书馆馆员的角色	徐纲红 杨 波(251)

◇信息学探讨与其他◇

浅论军事情报学	余诗武(255)
浅谈图书情报学术论文写作	叶春峰 艾润莲(258)
试论基层文献工作存在的主要问题及对策	张玉兰(262)
知识经济、知识管理与情报研究	张爱科(266)
论信息时代高校图书馆的现代化发展趋势	蔡 勤 郭 楠(269)
西部大开发与西部图书馆	滕春波(272)
二十一世纪图书馆的发展趋势	薛保延 王丽珍(275)
陕西省图书馆学研究中存在的问题与对策	李小鸽(277)
信息时代期刊工作的几点思考	吴俊仙(280)
陕西省图书馆的创建过程及其解放前的历史发展	杨昌俊(282)
图书馆综合文化功能的调查分析	金 科 孙金立 王 栋(286)
高校图书馆的职能新定位与“创新型”大学生信息素质教育	孙同月(289)
传统与现代相融,理论与实践共进	
——21 世纪图书馆发展战略研究	高坐仑(293)
信息市场当前存在的问题和对策	王丽珍 薛保延(296)
试论发展复合图书馆的基本策略	黄尧力 崔红雁(299)
新世纪信息产业如何抓住机遇与时俱进	鲍心清(302)
非合并院校多校区图书馆现状及发展之思考	张芳梅(307)
浅论陕西宋元以前的古方志	李文谨(311)
对图书馆业务档案管理现代化的思考	高文莉(315)
加快陕西省信息化建设的步伐	李 蓉(318)
高校图书馆读者统计的理论与实践	罗广玲(321)
应重视临床诊断信息的系统化研究	曹茂开(324)
网络环境下文献信息检索课教学改革的新思路	陈 泉(327)
浅谈“入世”后的我国信息产业及其对策	张雪绸(331)
加快西部图书馆发展的策略探讨	周应萍(334)
实践“三个代表”重要思想 积极发挥图书馆社会职能	张恩周(337)
迎西部大开发,创建西北情报信息网	张 峰 贺秀英(340)
论对著作权限制之合理使用制度	赵庆菊(343)

数字图书馆业务系统设计实践

王建平(西安建筑科技大学图书馆 西安 710055)

摘 要 本文从数字图书馆的概念出发,对业务系统的设计实现思想做了说明,在此提出了系统体系结构、系统组成等关键性问题,并分别提供了解决方法。另外就文献标引和著录模块这一核心部分,从功能分析和流程处理等方面提供了具体实现方法。

关键词 数字图书馆 图书馆业务 标引

图书馆业务系统是数字图书馆工程中的核心与重点。其目的是以计算机网络化的图书馆工作模式代替传统图书馆的封闭式手工作业方式,实现行业系统内各图书信息单位对文献资料的联合采购协调、联机联合编目加工、文献目录通报的自动排印和对连续出版物的有序管理,同时还可提供全面网络化的联机公共目录查询、联机文献借阅管理以及馆际互借服务。系统需要严格按照国际标准 UNIMAC 来统一规范各单位的文献采购和编目数据,减少各单位间文献的重复订购、重复加工,提高联合文献数据库的质量,降低用户查找文献的难度,提高文献的查全率、查准率和对一次文献的获取速度。如果实现得当,还可降低工作人员的劳动强度,加快文献传递速度,为读者提供更便捷的文献获取途径,实现信息资源更广泛的共享,从而最终实现图书馆业务的标准化、信息流程的实时化和文献处理工作的无纸化。

在研究数字图书馆概念的基础上,我们对业务系统的实现做了相关的设计尝试,以供数字图书馆建设参考。

1. 系统体系结构

系统的体系结构可设计如下:

采用这种结构,系统适用于任何与 Internet 连接的局域网络工作组或单机用户。另外,对不上网的单机用户,也能为其提供一些必要的功能模块。总之,其目标是尽可能多地发展用户,使系统文献数据积累日益增加,成为真正的文献数据中心。

为了大力发展用户,鼓励用户积极向中心服务器递送文献编目信息,同时又从中心服务器下载已有的文献编目信息。可以采取“上送有酬,下载低收费”的管理原则,使系统在最大范围内实现信息资源的共建共享。

2. 系统组成

2.1 系统功能

相应于传统图书馆业务,系统的作用应该充分体现在以下几个方面:

- 避免不必要的文献重复订购,保证资金使用更趋合理;
- 避免文献重复编目方式,降低劳动强度;
- 按国际标准统一规范各单位的文献编目数据,提高科技文献联合数据库的质量;
- 使图书馆各业务环节工作更趋合理,更为科学,并提高自动化程度;

- 使图书馆文献管理更加透明,提高文献收集的针对性,加快文献的传递速度;
- 降低用户查找文献的难度,提供更快捷的文献获取途径,使信息资源得到更广泛的共享和利用。

根据以上要求,并且为了达到应用于数字图书馆的目的,系统的主要功能可以概括为:

- 1)实现多馆间联机文献采购协调,以节省采购经费;
- 2)实现多馆间联机联合文献编目加工,以避免重复劳动;
- 3)实现全面网络化的联机公共文献目录查询和馆际互借服务,以实现资源共享;
- 4)实现图书馆业务标准化、信息流程的实时化和文献处理工作的无纸化。

2.2 系统结构

整个系统的工作流程主要包括文献搜集/订购、帐目总登、文献登录、文献标引、目录报道、文献检索、文献入藏、流通管理以及网络数据库服务器管理等模块;文献处理的种类有中文文献、外文文献、中文期刊、西文期刊、密级文献和成套文献。而每种文献都有各自不同的处理流程,尤其是连续出版物有其独特的订购和管理流程。因此,系统可从8个方面实现,这些方面既相互关联,又可独立使用。

- 文献搜集/订购模块
- 文献登录模块
- 文献著录和标引模块
- 文献目录通报出版和报道模块
- 全文检索模块
- 文献借阅流通模块
- 连续出版物管理模块
- 系统服务和管理模块

3. 功能实现

一个完整的图书馆业务系统,其各个模块的实现都有特殊性和技术难点。我们在文献标引和著录方面做了一定的研究,下面就以此模块为例加以分析。

3.1 功能分析

文献标引和著录模块的主要功能是形成符合规范的书目数据,并能将书目数据置于广域网的环境下,实现书目信息资源的共建共享。文献标引和著录模块的设计主要从使用的功能要求和数据的流程进行考虑。

相应于传统的编目方法,即使是做到了计算机化,但原有的编目部分常常是在局域网环境下进行,而我们所完成的文献标引和著录模块则考虑到了广域网环境下书目记录的编制和共建共享,所以其基本控制有所不同。另外,从数据项要求方面来说,传统编目部分的字段通常要求比较简单,字段基本是固定的,并在相当长的时间内不会有大的改变,因而界面的数据库设计比较简单,使用的是一个数据库字段对应一个书目记录的字段。而这里则相当复杂,需要使用 MARC 作为编辑和存储的主要数据格式。在此主要考虑以下几个因素:

- 1)书目数据是变长的,如题名、摘要等;
- 2)数据项是重复出现的,并且重复的次数不固定;
- 3)数据项的出现是随意的;

4) 记录间的联系方式是复杂的,如连接、合并等;

5) 文字方式是复杂的。

这些问题,使用传统的字段对应方式显然不能很好地解决。所以书目数据库使用书目数据实体和检索用索引相分离的方法。一个数据表中存储 MARC 数据,另一个表以对应方式提供简单检索,两个表之间以记录号相联系。在编目检索时,实际上是利用 MARC 检索用表;在向 MARC 数据表中写入记录时,同时修改 MARC 检索用表数据,这两个表的数据时刻保持一致。

3.2 功能内容

标引和著录模块在功能上需要实现如下内容:

- 编目前文献验收,订购部门单据核对
- 编目后文献入藏,打印入藏清单
- 检索书目数据
- 检索订购数据
- 下载书目数据
- 下载数据,由于订购数据以字段形式存放,应该能够组合成 MARC 文本
- 编辑下载后的数据
- 新增、修改书目记录
- 对字段、子字段定义
- 定义工作单,并为工作单提供默认值
- 提供编目辅助信息
- 提供计算机辅助标引系统

3.3 流程设计

文献标引和著录模块的流程主要从两个方面考虑,一是文献实物处理的流程,二是软件系统处理数据时的流程。前者是基础,后者根据前者设计,原则是保持前者顺利、良好、高效运行。

从文献实物处理流程上看,当文献处理人员接收到文献实物后,应当核对单据和文献实物,这可以由处理人员手工完成。同时还需要把实体文献对应的文献号(条码)对应到流通系统中。

文献标引时,首先查找文献是否已经标引过。如果文献已经标引,直接下载书目数据到客户机上进行标引;如果文献没有标引,则需要检索订购表,下载订购记录并形成一条新增的书目记录。

编辑书目记录,对书目记录字段进行增、删、改并保存,提供有效的计算机辅助标引系统。编辑完书目记录后,需要入藏文献实体,这时应打印出入藏清单,并将数据发送到其它子系统。

以上是文献实物处理流程,软件也应有与之相适应的基本功能。不同的是软件需要在安装后第一次使用前进行一些设置。

参考文献

1. 张义霞.二十一世纪图书馆的发展方向——数字图书馆.中国研究,2001(9)
2. 张道义.数字图书馆构成要素探讨.图书馆论坛,2002(4)
3. 傅旭文.图书馆数字化建设的探讨.图书馆建设,2002(4)
4. 牛雅杰等.论图书馆数字化面临的十大难题.情报杂志,2002(3)

磁盘存储技术在电子阅览室中的应用

郭梅(西安电力高等专科学校图书馆 西安 710032)

摘要 在分析现代电子阅览室的重要性的基础上,介绍了两种主要的存储设备,即光盘镜像服务器和磁盘阵列,并且分析了主要的磁盘存储技术,包括:NAS和能够有效提高计算机传递速度和数据可靠性的廉价冗余磁盘阵列(RAID)。详细分析了RAID的四个常用级别即RAID0、RAID1、RAID3、RAID5,并介绍了如何保证磁盘阵列的安全性,以及如何计算磁盘阵列中磁盘的容量。

关键词 电子阅览室 磁盘阵列 RAID 光盘镜像服务器 NAS

1. 绪论

随着文献数字化进程的加快,图书馆已走出了原有的封闭、落后、烦琐的模式,馆藏中光盘文献资源的比重越来越大。许多大型的数据库,如基于Web页面检索的中国科技期刊全文数据库,重庆维普中英文科技期刊数据库,万方摘要数据库等,都涌现出来;另一方面,图书馆也不再是仅仅为读者提供书籍和资料,更重要的是要能够将文字、图像、声音以最接近真迹的生动活泼的形态提供给读者,并使其充分流通起来。那么如何充分发挥光盘文献资源的作用,提高其利用率,成为图书馆界要探索的新课题。而在这一课题中,首先要解决的就是大容量存储的问题。图书馆存储信息的度量单位由KB、MB到GB、TB甚至PB,传统的单个硬盘已经无法满足其大容量数据存储的要求,因此成功地建立一个电子阅览室,除了需要建立一个安全稳定的网络机房之外,最重要的莫过于对这一大型存储设备的选择了。目前典型的存储设备有基于NAS的光盘镜像服务器和使用RAID技术的磁盘阵列,下面就分别对这两种设备进行探讨。

2. 光盘镜像服务器

光盘镜像服务器是一种将硬盘高速缓存技术和网络服务器技术相结合,专为光盘网络共享而设计的NAS设备。光盘镜像服务器本身没有通用服务器那样复杂的操作系统和硬件连接,只完成光盘镜像服务器硬盘数据和客户机之间的数据传送,使客户机能以硬盘的访问速度来共享CD-ROM光盘上的信息资源,消除了CD-ROM驱动器瓶颈问题,极大地改善了光盘网络共享的性能。既然光盘镜像服务器是一种NAS设备,那么什么是NAS呢?

2.1 什么是NAS?

NAS是一种具有很大存储容量的设备,它通过集线器或交换机直接连在网络上,使磁盘空间的扩展如同在网络上添加打印机一样方便。NAS设备的物理位置很灵活,既可以放置在数据中心的工作组内,也可以放在其它地点,通过物理链路与网络连接起来。NAS设备集成了存储器件(例如磁盘驱动器阵列,CD或DVD驱动器,磁带驱动器或可移动的存储介质)和简易服务器,具有RAID功能和完全的文件服务器功能,集成在NAS设备中的简易服务器可以将有关存储的功能与应用服务器执行的其他功能分隔开,允许用户无需应用服务器的干预,

就可以在网络上存取数据。这样既可减少 CPU 的开销,也能显著改善性能。虽然目前市场上有许多存储设备,如 SAN 大型磁带库、磁盘阵列、光盘库等,但其昂贵的价格和复杂的操作是资金和技术实力有限的中小型图书馆不能接受的。NAS 正是这种需要的产物,相关技术的成熟是 NAS 产品快速发展的重要因素,尤其是 IDE 硬盘技术的飞速发展和 IDE RAID 技术的引入,可以说是 NAS 产品发展的里程碑。

2.2 光盘镜像服务器 CDH

将 NAS 设备连接到网络上非常方便,NAS 设备提供 RJ-45 接口和独立的 IP 地址,可以将其直接挂接在主干网的交换机或其它局域网的 HUB 上,通过简单的设备就可以在网络上应用,完全可以即插即用。同时 NAS 设备允许用户在网络上存取数据,有效改善了网络的性能。NAS 设备还有动态的扩展能力,即使需要添加新的 NAS 设备,也无需停顿整个网络服务,从而保证数据流畅存储,十分适合 7×24 小时的全面联机服务体系结构。

许多学校都采用了典型的 NAS 应用,把 NAS 应用于核心交换机之上,不仅可以使电子阅览室调用其内部视频文件,同时作为图书馆的电子图书设备,可以储存大量电子图书资料。光盘镜像服务器 CDH 作为电子图书馆 NAS 系统中的主要设备,其工作原理是:CDH 中既有普通的 CD ROM 光驱,又包括一些大容量的硬盘,用户可以通过塔中的 CD 光驱或网络中任一工作站中的 CD 光驱,分别将各种数据库光盘自动压缩或镜像到 CDH 中的硬盘上,然后可以取出 CD 光驱中的数据库光盘;此时 CDH 服务器中的硬盘已包含以光盘卷为单位的很多张光盘数据库的镜像,网上用户共享硬盘中的“光盘”就如同共享 CDH 服务器上的 CD 光驱中的光盘数据库一样,不同之处是速度提高了 60 倍,在线同时共享的光盘数量也大大增加。光盘镜像服务器(CDH 型)最大能容纳 8 块 IDE 硬盘,通过内部管理软件可将硬盘划分成文件镜像和文档共享两种模式,最多可镜像 1000 片容量不限的光盘,可将档案资料储存在 255 个容量不限的文件目录下。它是市场上解决网络资源存取产品中容量较大的设备。由于采用 IDE 接口和硬盘,大大降低了购置成本和维护成本,并且 CDH 自身就是一台服务器,安装极其简便,只要接上电源和网络,在 LCD 控制面板上设定好 IP 地址即可,不需要安装任何驱动程序和软件,摆脱了光盘柜类产品需要另接一台服务器作为管理设备的繁琐模式。在安全性方面,使用 RAID 0、1、3、5,并可接刻录机和磁带机,因此提供了优良的资料安全备份功能。在资料读取和写入的安全控制上,可设定管理者密码与使用者对资料的使用权,管理者还可将用户资料直接建立在文件镜像服务器上,镜像在服务器上的光盘资料不能更改,有效防止电脑黑客和病毒的侵害。此外,大多数的光盘镜像服务器对资料共享都做了特别设计:读取头移动最佳化可有效降低光盘的平均存取时间;以网络连线作预测的预读暂存可针对每个网络使用者,在读取需求送到服务器之前先把资料准备在快速内存之中,使多人可同时读取一片光盘而不会有断读或跳格的现象。

2.3 光盘镜像服务器的特点

这种以使用 NAS 系统 IDE 硬盘为主要特点的光盘镜像服务器,支持 TCP/IP、FTP 等网络协议及多种操作系统如 Unix Dos、Win 95、Win 98、Win 2000、Win NT/ME 等。这种存储设备在速度上低于使用 SCSI 硬盘的磁盘阵列,在安全性上也不及使用真正硬 RAID 的磁盘阵列,但在价格上却是很有优势的,它的价格是相同磁盘阵列的四分之一。同以前广泛被使用的光盘塔相比较,光盘镜像服务器也有很强的优势,光盘镜像服务器将光盘的信息存储和读取功能分离,凭借硬盘的高速存储能力来共享光盘信息资源。因此,光盘镜像服务器的访问速度要

比光盘库或光盘塔快几十倍,光盘镜像服务器在容量和速度等性能指标方面均超过光盘库和光盘塔。但是其单位容量成本却大大低于光盘库和光盘塔,光盘镜像服务器给电子阅览室等类似用户提供了一种性价比很高的光盘网络共享解决方案。

在使用中光盘镜像服务器的存储部分很象一个储物的大桶,文件是以 one by one 的形式存入,无法在存入时就指定好存储的目录(以卷的形式存在),但 econtent 管理软件解决了这一问题。econtent 作为 CDH 的管理软件,通过其独有的管理方式,把镜像服务器的文件资料及镜像资料进行归档,使用户可以方便地从自有的文件夹中找出有用的资料,同时管理员又可进行权限管理。

虽然光盘镜像服务器有许多的优势,但值得注意的是,光盘镜像服务器中使用的是软件 RAID。软件 RAID 的阵列操作和管理功能由运行于主机 CPU 上的阵列软件控制,阵列软件会与其它应用程序争夺 CPU 时间和内存,因此软件 RAID 降低了服务器的整体性能。

3. 磁盘阵列

3.1 什么是磁盘阵列

磁盘阵列是由一个硬盘控制器来控制多个硬盘的相互连接,使多个硬盘的读写同步,减少错误,增加效率和可靠度的技术,是一种大容量的存储设备。所谓大容量在磁盘阵列的概念中,并不是指市场上所谓的 80GB、90GB、100GB、180GB 的单个硬盘,硬盘容量较大叫大容量,而是指将这些单个硬盘透过 RAID 的技术,按 RAID LEVEL 组合成更大容量的硬盘,从 100GB 到 5000GB(5TB)。依单个硬盘的个数容量,可组合成不同超大容量的磁盘阵列系统。从主机端来看,磁盘阵列是一个超大硬盘,同样能分区,而它的功能,则要看你选择的 RAID Level 的级别。下面我们从三个方面来探讨磁盘阵列。

3.2 RAID 技术

RAID 为 Redundant Arrays of Independent Disks 的简称,中文意思是廉价冗余磁盘阵列,是磁盘阵列在技术上实现的理论标准,其目的在于减少错误,提高存储系统的性能与可靠度。目前 RAID 已广泛地被使用,RAID 也有多个级别,最常见的是 0、1、3、5 四个级别,其他还有 6、7、10、30、50 等。RAID 技术是 1987 年由加州大学伯克利分校提出的。最初的研制目的是为了组合小的廉价磁盘来代替大的昂贵磁盘,以降低费用,同时希望采用冗余信息的减少,使得磁盘失效时不会使对数据的访问受损失,从而开发出一定水平的数据保护技术。用 RAID 技术可以把速度提高到单个硬盘驱动器的 400%,磁盘阵列把多个硬盘驱动器连接在一起协同工作,大大提高了速度,同时把硬盘系统的可靠性提高到接近无错的境界。这些“容错”系统速度极快,同时可靠性极高,下边我们先仔细了解一下常见的 RAID 种类。

3.2.1 RAID 0

它是将多块硬盘(至少 2 块)连接在一起而组成一个容量更大的存储设备,将数据以子系统规定的段为单位写入多个磁盘中,例如数据段 1 写入硬盘 0,段 2 写入硬盘 1,段 3 写入硬盘 2 等。当数据写完最后一个硬盘时,它就重新从盘 0 的下一可用段开始写入,写数据的全过程按此重复直到数据写完。段由块组成,而块又由字节组成。由于硬盘分段的方法,是把数据立即写入(读出)多个硬盘,因此它的速度比较快。实际上,数据库的传输是顺序的,但多个读(或写)操作则可以重迭进行,也就是,正当段 1 在写入驱动器 0 时,段 2 写入驱动器 1 的操作也开始了,而当段 2 尚在写入驱动器 1 时,所写数据已送驱动器 2,如此类推。在同一时刻有几个

盘同时写数据,这就大大提高了 I/O 速度。在所有 RAID 级别中,RAID 0 的速度是较高的,但是 RAID 0 没有冗余功能和错误修复能力,只要有一块磁盘出现故障,整个 RAID 系统的数据都将丢失,无法补救,这种缺憾无疑是严重的。虽然 RAID 0 可以提供更高的空间和更好的性能,但是整个系统是非常不可靠的,可用的磁盘数为 n , (阵列中共有 n 个硬盘)。

3.2.2 RAID 1

又称为磁盘镜像,最少需要 2 块硬盘。它是将一个硬盘中的所有数据拷贝到第 2 个硬盘中,使两个硬盘互为镜像,系统可以从一组镜像盘中的任何一个磁盘读取数据。在 RAID 1 中,任何一块硬盘的损坏都不会影响系统的运行,系统一旦发现坏的硬盘,便会立即去寻找另一块镜像盘,RAID 1 可以在一半数量的硬盘出现问题时继续工作。在 RAID 1 中磁盘的使用率是 $N/2$,这就大大增加了系统的成本,它也是所有 RAID 种类磁盘利用率最低的一种。

3.2.3 RAID 3

它是采用磁盘分段加固定校验的方式,阵列中的 N 块硬盘,将有一块硬盘被用于专门保存校验信息,而数据则像 RAID 1 中那样均匀的存放在其它的 $n-1$ 块硬盘中,并能输入/输出。从表面上看 RAID 3 拥有 RAID 1 的速度,又避免了 RAID 1 中无纠错能力的缺点,但事实却并非如此。在 RAID 3 中系统会因校验而受到影响,当一块硬盘失效时,该硬盘上的所有数据必须使用校验信息重新建立。当我们更换了已损坏的磁盘后,系统也必须一个数据段一个数据段地更换坏盘中数据。因此这种维修工作最好放在 RAID 系统空闲时进行。鉴于这种 RAID 的特殊性,RAID 3 更适合用于多读出操作、少写入操作的环境,它的磁盘可用的为 $n-1$ 。

3.2.4 RAID 5

它是对 RAID 3 的改进,与 RAID 3 最大的区别是不再将校验信息写在一块磁盘上,而是均匀地存放在系统中的每一块磁盘上。它大大改善了 RAID 3 中可能出现的 I/O 瓶颈效应,提高了子系统的事务处理能力。RAID 5 是目前被广泛使用的一种 RAID 技术,它将高性能、低成本、高安全性集于一身。对硬盘操作来说,RAID 5 的写性能比不上直接硬盘分段(指没有校验信息的 RAID 0),因为产生或存储奇偶码需要一些额外操作。例如,在修改一个硬盘上的数据时,其它盘上对应段的数据(即使是无关的数据)也要读入主机,以便产生必要的奇偶信息。奇偶段产生后,要将更新的数据段和奇偶段写入硬盘,因此,虽然 RAID 5 比 RAID 0 优越,但就写性能来说,RAID 5 不如 RAID 0。另一个要注意的问题是,采用 RAID 5 时,对于 5 个硬盘的数组,有大约 20% 的硬盘空间用于存放奇偶码,而 10 个硬盘的数组只有 10% 的空间存放奇偶码。在可用空间总的格式化空间的意义上来说,硬盘系统中的硬盘越多该系统就越省钱。

3.3 磁盘阵列的安全性

使用磁盘阵列的好处,在于数据的安全、存取的速度及超大的存储容量,而数据的安全则主要看该阵列是否有环境监控器针对温度、电压、电源散热风扇、硬盘状态等进行监控。在实际工作中,一套系统中各个部件往往是环环相扣的,一个部件坏了经过一段时间不处理则会引起其它部件的损坏。最典型的例子便是一个 4 美元的风扇坏了,因为散热不好,引起价值 400 ~ 500 美元的 CPU 温度过高而烧坏,服务器死机了,我们才发现,可是为时已晚。问题的关键是我们没有在风扇工作不稳定时就及时地更换风扇从而避免巨大的损失,因此环境监控器是防治不稳而不是用于发现什么坏了。环境监控器可将温度的值显示在面板上,随时了解温度的变化,另外再将报警温度设定,如温度高达警界点,磁盘阵列就会自动报警。

3.4 如何计算存储容量

从主机端看磁盘阵列是一个大容量硬盘,从 Host 端来看磁盘阵列是看到逻辑的 Array 1、Array 2、Array 3、Array 4。在 Array 中可以将 Array 分成多个逻辑区(LUN),就和在 SCSI 盘上分 LUN 一样,每个 Array 由几个硬盘组成是有限制的,不同品牌有不同的限制。我们可以用 Array 1 建 RAID 0,Array 2 建立 RAID 5,Array 3 建立 RAID 1,Array 4 可空白不用。

下面用例子来说明如何计算磁盘存储容量。

例 我们用 8 个硬盘来规划硬盘阵列,每个盘容量为 80GB。

Array 1:用 2 个盘做 RAID Level 1(镜像)

Array 2:其它 6 个盘做 RAID Level 5。则可用容量为:

Array1 的可用容量为: $(2 \times 80\text{GB})/2 = 80\text{GB}$

Array2 的可用容量为: $(6 - 1) \times 80\text{GB} = 400\text{GB}$

全部可用容量为:Array1 + Array2 = 80GB + 400GB = 480GB

分析:原来 8 个 80GB 的硬盘的总容量为 640GB,但是实际可用容量为 480GB,少了 160GB,也就少了两个硬盘。原因是 Array 1 做 RAID Level 1(镜像)及 Array 2 做 RAID Level 5, Array 1 及 Array 2 各浪费了一个硬盘所致。因此在规划磁盘阵列的容量时,我们既要考虑实用性又要考虑经济性。在这个例子中,Array 1 是多此一举,因为 Array 1 做 RAID Level 是因为数据要有备份,但 RAID 5 本身就已有容错的功能,不需考虑备份,实用性这一面已不需考虑,只需考虑其经济性。我们将 Array 1 和 Array 2 合二为一为一个 Array 1。Array 1 的可用容量为 $(8 - 1) \times 80\text{GB} = 560\text{GB}$ 。这样规划只浪费了一个盘的空间,实际可用容量为 560GB,比第一方案又多出了 80GB。

光盘镜像服务器中光盘数据按介质原始记录格式读取存放在硬盘,成为一个镜像文件,所有光盘的数据实质组成一个网络文件服务器。在操作系统基本的网络协议支持下,通过设备驱动模块对用户访问的镜像文件虚拟成光驱,其上层的应用管理层在文件共享协议支持下实现光盘数据共享,使具有不同权限的用户访问不同的文献资源,而且局域网和 Internet 用户都可通过 WEB 浏览器来访问镜像服务器上的光盘文献资源。在实际工作中,我们发现光盘镜像服务器最大的优点就是维护使用方便,但缺点是不能运行客户的应用程序,升级扩展性差。

短短的几年时间,图书馆中的大容量存储设备已由最初的光盘库、光盘塔发展到了光盘镜像服务器、磁盘阵列,已由最初仅仅是以解决大容量存储为目的发展到今天要求速度快,安全性高,而新的更先进的存储设备又在相继产生。相信先进的存储设备必将把电子阅览室带向一个更广阔的天地。

参考文献

1. 普桦豪威公司. 磁盘阵列 双机容错白皮书
2. 刘宇. 磁盘阵列技术在图书馆的应用. 现代信息技术, 2001(2)
3. 刘维志. 初探磁盘阵列. 网络与信息, 2001(11)

网络数据库的质量评价

张幸芝(西安交通大学图书馆 西安 710061)

宓 欣(西安交通大学研究生院 西安 710049)

摘 要 本文主要从用户的角度出发,对目前 Internet/Intranet 上的网络数据库进行质量分析和评价,得出现有网络数据库在数据、检索功能、操作、维护及运行等几个方面的优点和不足之处,使用户有一个选择网络数据库的衡量尺度。同时针对存在的问题,提出了相应的建议和对策,这对日后网络数据库检索系统软件的完善有一定的借鉴参考价值。

关键词 网络数据库 检索途径 质量评价

目前,信息通过网络来传播,已经被视为最快捷的方式。不同的行业和专业对信息的需求不同,网络数据库可以有针对性地将信息分类建库,形成不同的专业数据库和公共信息数据库,从而满足不同的需求。不仅如此,网络数据库还具有数据资源丰富、检索方式多样、传播和携带便捷等优势,因此越来越多的用户开始倾向于使用网络数据库来查找自己所需的资料和信息。那么,究竟我们通常所使用的数据库质量如何,笔者认为应该从以下几个方面进行评价。

1. 数据质量

1.1 数据来源

数据的来源直接影响着用户所检索到的信息是否准确、全面。专业文献数据库在收录文献资料时要求比较严格,一般需要和文献出版商签订合同,以保证文献资料的合法性、权威性、完整性和时效性,使用户能在第一时间查到最完整、准确的资料。对于公共信息数据库,例如:网上搜索引擎(常用的有 Google、千百度等),大部分信息是从 Internet 上收集到的,此类数据库是为了给人们一个查找信息的入口,至于信息的准确性和可靠性需要用户自己去判断。

1.2 收录范围

评价一个专业文献数据库的收录范围,主要是看数据库中所收录文献的专业性和回溯年限,即最大限度地收录某一专业领域的文献资料,强调专业性的同时,还要考虑学科间的渗透关系,收录交叉学科文献资料。除此以外,还要看数据库中出自重点刊物的文献所占的比重以及特色文献收录,如:科技成果、发明专利、国家标准、技术开发项目、科技立项、行业动态等。公共信息数据库主要是看信息的覆盖面,即是否包含了工作、生活等各个方面的信息,强调数据的综合性。

1.3 数据记录的字段设计

数据库中的数据是以记录为单位的,每条记录由字段组成,如:篇名、作者、摘要、关键词等多个字段组成一条文献记录。记录的字段并不是越多越好,而是根据用户检索的需求合理设计,在建库之前,最好能够在数据库的使用对象中作一次调查。

2. 数据库所具备的功能

数据库的检索功能是否强大,直接影响着检索效率的高低,其中检索途径(策略)和结果处理是网络数据库所具备的基本功能。公共信息数据库通常采用单项模糊检索和二次检索,其检索结果事实上又是一个新的查找入口,因此就不存在诸如存盘、打印、排序等结果处理之说了。采用 Client—Server 方式运行的专业文献数据库,如中国生物医学文献数据库(CBM disc),所有的检索功能都包含在客户端(Client)应用程序中,此类检索系统在检索途径、策略及结果处理等方面设计的比较完善,值得借鉴参考。

然而,大多数基于浏览器—Web 服务器方式运行的网络数据库,在检索功能方面尚欠成熟。下面就现有的专业文献数据库功能加以讨论评价。

2.1 检索途径

检索途径一般分为单项、组合检索两种,大体上网络数据库都具备这些功能,但在细节方面还存在一些差异,主要体现在单项检索中。在检索时,不难发现多数网络数据库都采用模糊匹配,这样的确可以使检索结果最多,保证了查全率,但要求用户必须耐心地在数百条甚至上千条记录中逐个挑选,再加上网络速度的限制,使用户费时费力、头疼不已。如果能考虑到用户的需求差异,再多设计几个匹配选项,如:完全匹配、模糊、中间相同、前(后)端相同等,会更有利于用户在检索过程中的筛选,节省查找资料的时间。

另外,大多数网络数据库中的关键词是由原始作者提供的,缺乏专业标引人员人工干预,在此基础上建立的主题检索,其准确率必然会受到影响^[1]。关键词检索是单项检索中使用率较高的一种检索途径,因此非常有必要增补关键词将其完善,提高检索结果的准确率。组合检索是专业文献数据库中必不可少的一种检索方式,通过多个字段的共同限制检索资料,既可以精确检索结果,也增加了检索的灵活性。其中组合字段是用户使用率较高的字段,如篇名、作者、出版日期等,组合字段的数目可以根据数据库容量大小来选择,组合关系通常包括与(AND)、或(OR)、非(NOT)三种,不过选用“与”关系的较多,“或”和“非”比较少见。各数据库的组合过程、方法和组合字段的数目略有差别,有些将主要字段的组合设计在一个页面内,称其为高级检索;有些则在每个单项检索页面内,都设计一个“组合”按钮,以备用户随时调用。总之,无论采用哪种方式,都要以方便用户检索为前提。

2.2 检索结果的处理

在使用了大小十几个网络数据库之后,发现显示、存盘、打印是常见的检索结果处理方式,而且已经做得比较完善,用户还可以通过记录标识有选择的处理检索结果,但排序则不常见到。通常用户都希望看到最新发表的文献资料,在没有排序功能的系统中,如果检索结果还是按照时间先后顺序排列,用户自行翻页所造成的不便可想而知。

3. 用户操作界面的灵活性和便利性

3.1 界面的合理布局

界面是网络数据库系统提供给用户操作的一个平台,除了追求视觉效果以外,整个检索界面的布局还要非常合理才行。所谓合理,简单地说就是将用户常用的操作和帮助设计在便于使用的地方,界面之间的切换要有明显的标识,以提高操作的灵活性。