

联机编目与数字图书馆

胡广翔 主编



 科学技术文献出版社

2003. 8

中国社会科学院文献信息中心重点研究课题

“全院联机编目研究与实践”研究成果

联机编目与数字图书馆

顾问 袁名敦

主编 胡广翔

北京·科技文献出版社

2003年7月

图书在版编目 (CIP) 数据

联机编目与数字图书馆/胡广翔主编. —北京: 科学技术文献出版社, 2003.7

ISBN 7-5023-4358-X

I. 联... II. 胡... III. ①机器可读目录—编目—文集②数字图书馆—文集 IV. ①G254.362-53②G250.76-53

中国版本图书馆 CIP 数据核字 (2003) 第 050819 号

责任编辑: 科文

联机编目与数字图书馆

科学技术文献出版社

北京市复兴路 15 号 (中央电视台西侧) /100038

北京市后沙峪印刷厂

880×1230 毫米 1/32 • 10 印张 • 252 千字

2003 年 7 月第 1 版 2003 年 7 月第 1 次印刷

印数: 3000 册 定价: 48 元

ISBN 7-5023-4358-X/G • 963

编 者 的 话

中国社会科学院联机编目研究与实践课题是我院文献信息中心的重点研究课题。按照该课题的计划,在完成了全院实时的联机联合编目系统和与之相关的课题研究报告之后,还要编辑、出版一本由本课题研究人员撰写的有关联机编目和数字图书馆内容的文集。该文集的内容既包括学术理论文章,也包括全面介绍国内外有关联机编目和数字图书馆比较研究和建设现状的文章,以及收集比较齐全的有关数字图书馆的各种参考文献的书目,这些参考书目既包括用中英文出版的印刷本,也包括网上刊行的电子文献和数据库等。因为联机联合编目系统是建设数字图书馆的重要基础,而未来的联机联合编目业务也应融入未来数字图书馆的总体建设工程之中,所以我们才有了出一本以这样两个主题为内容的文集的想法。而且,该文集的侧重点还是在数字图书馆方面,因为,即使联机联合编目搞得再好,毕竟我们最终的目标是要建成符合信息时代和知识经济时代要求的、超越时间和空间限制的、实体图书馆与虚拟图书馆相结合的数字图书馆。

我们出版此文集的目的主要有两个,一是对前一阶段的研究做一个概括和总结;二是为关心、了解联机编目和数字图书馆有关状况的人员提供一些最基本的参考。由于我们对这些问题的研究还很欠缺,水平也很有限,谬误之处敬请各位专家学者不吝指教。

联机编目与数字图书馆

目 次

一、中国社会科学院联机联合编目系统——方案设计及实施	1
二、中国社会科学院联机编目研究与实践	19
三、实现联机联合编目的基础与条件	56
四、数字化图书馆技术	103
五、中国数字图书馆研究综述	115
六、国外数字图书馆	137
七、韩国数字图书馆研究	157
八、中文数字图书馆参考文献	225
九、数字图书馆英文参考书目	265
十、网上数字图书馆文献索引	306

中国社会科学院联机联合编目系统

——方案设计及实施

● 胡广翔 庞 萱

[摘要] 本文结合时代背景综合分析了联机联合编目作业系统的工作模式，论述了根据实际需要规划设计的中国社会科学院联机联合编目系统总体方案，介绍了中国社会科学院联机联合书目数据库的开发、使用及组织管理方法。

[关键词] 联机联合编目 系统设计 数据库开发

目 次

- 一、历史回顾与时代背景
- 二、系统方案选择
- 三、系统设计
- 四、管理制度和人员培训
- 五、成果回顾与展望

一、历史回顾与时代背景

文献编目在我国有着悠久的历史，古代目录学家创造了许多宝贵的记录文献的方法，积累了丰富的组织藏书的经验。在我国目录学史上虽然没有明确提到联机编目，但我国古代的目录工作已含有联合编目的萌芽。从我国最早的总书目《七略》、《汉书·艺文志》、《隋书·经籍志》、《唐书·艺文志》到郑樵的《艺文略》，均有编制全国综合性的古今图书系统目录的意图。由于历史条件的限制，中国古代官私藏书一向不开放，所以古代所编的总书目都未能揭示各书的藏所，这与联合编目的宗旨还不相符。十七世纪明代僧人智旭编制的《阅藏知津》就注明了收藏佛藏的方位，它的编制方法已体现出共享的思想。在三十年代北平图书馆曾经开展铅印卡片工作，这可说是我国集中编目的开端。新中国成立后，1957年11月正式成立了全国图书联合目录编辑组，附设在北京图书馆内。它的任务是：（1）了解、调查全国各图书馆的馆藏和编目情况；（2）制订联合目录编辑计划；（3）起草联合目录编目条例；（4）加强各馆有关联合目录工作的联系、布置、检查和督促工作；综合各馆书目，做最后的编排、校定、出版等工作。从一九五七年～一九六六年的九年间，共出版全国性和地区性的联合目录300余种。这一时期，全国图书联合目录编辑组出版的《全国中文期刊联合目录》、《中国古农书联合目录》、《中医图书联合目录》、《全国西文期刊联合目录》、《全国俄文期刊联合目录》、《全国日文期刊联合目录》等和上海图书馆出版的《中国丛书综录》都成为图书馆界、学术界必备的重要参考书目。

如今，人类社会已经跨入了21世纪，信息高速公路的建设和不断完善给社会带来了巨大的经济效益，使计算机信息网络成为世界上覆盖面最广、规模最大、信息资源最丰富的信息网络，它在全球基础设施中扮演着尤为重要的角色。世界上各个国家纷纷投入了大量资金开展这一工程，其主要原因在于：当四通八达的

通信网络建成之后，人类的一切文明成果都将真正地实现共享。传统图书馆的管理与工作方式随着信息时代的到来已经不能适应现代图书馆自动化系统的要求。图书馆自动化系统的出现及其发展使得图书馆的工作发生了根本的转变。围绕着图书馆自动化系统，图书馆各项工作的开展是相互依赖和协同发展的。近年来计算机网络的发展，给图书馆工作又注入了新的内容，这就是如何利用网络系统实现文献信息资源共享。

中国社会科学院有相当丰富的文献信息资源，但由于历史的原因，社科院隶属的研究所遍布北京城的东、西、南、北，地理位置相当分散，因而造成了文献信息资源的分散。再加之管理手段一直比较落后，提供高效率的服务变得十分困难，不能有效地充分发挥这些宝贵专业资源的作用。因此，尽快建成全院联合书目数据库已迫在眉睫。

1999 年社科院局域网的建成并开通为我们开发建设联机联合编目系统奠定了网络环境上的基础，创造了良好的硬件条件。同时，近年来图书馆自动化软件也已逐渐成熟。这些都使实现全院文献资源共享，减少重复和盲目的投入，彻底摆脱传统落后的工作方式，逐步实现业务工作的全面自动化，使图书馆工作向标准化、规范化、合作化的方向发展成为可能。

二、系统方案选择

纵观国内外成功的联机联合编目系统，实现方案一般分为三种：集中式设计、分布式设计和综合式设计。

(1)集中式设计方案：各馆保持自己完整的文献库，并与一个集中管理的中心数据库相连，建立实时的数据传送机制，中心数据库存放多馆联合目录、规范文档、主题词表等集中管理数据。具体操作步骤为，单馆编目时，首先查询联机联合目录库，如果在中心库中查到相关记录，则加入本馆馆藏字段并将该记录远程

套录到本馆的库中。如果中心库没有查到相关记录,则在中心库中抢注并在本馆系统中加工形成编目数据,存入本系统的同时也传送到远程的中心数据库中。

优点:设备简单,整齐划一,操作简单和方便,管理也简单方便,实时性好掌握。

缺点:过于集中易出瓶颈,即很多终端同时向主机连接(请求)时,主机速度再快也应付不了,终端只能等待(分区也忙不过来)。主机任务繁重,对终端的要求比较苛刻。

(2)分布式设计方案:各馆都保持自己完整的数据库和规范文档等,利用分布式数据库技术,建立完全的分布式数据库管理系统。在物理上不存在中心数据库或中心索引库,只存在逻辑上为全局统一模式的“单一数据库”。编目员在编目时,无需考虑到复杂的网络环境,所有的操作对编目员来说都是透明的。

优点:主要是克服集中式的过于集中的缺点,能较好地解决瓶颈问题。

缺点:程序比集中式复杂,由于复杂,出错的可能性就加大。另外,各主机尽可能是同种机为好。

分布式对计算机软硬件的要求比集中式高,因而只有在计算机发展到相当发达的程度下才有可能实现。

(3)综合式设计方案:这一方案考虑了集中式方案数据冗余量大和中心库庞大的缺点和分布式设计方案还未达到实用的问题,选择了中间道路。各馆保持各自的数据库,中心数据库存放各馆数据索引汇集成的总索引和规范文档。单馆编目时,查询目标都指向总索引,如果检到相关记录并且属于他馆的记录,则将该记录套录到本系统,同时修改记录并加馆藏字段;如果没有检到相关记录则新增编目记录,同时系统将索引传送到总索引库。

以上三种方案都有自己优势,我们根据各自的发展环境采用合适的方案应是明智的选择。第二种方案是将来的发展方向。

第三种方案适用于跨系统大型图书馆系统间的联机合作编目,不需要重复投资,利用现有设备即可实现联机合作编目。第一种方案适合于层次性的图书馆系统,即由本系统中高层次图书馆维护中心数据库,带动第二层图书馆实现联机联合编目。就中国社会科学院来说,中国社会科学院文献信息中心自动化工作开展得比较好,书目数据资源比较丰富,具有建立和管理集中式中心数据库的条件和能力。而社科院其它研究所技术力量相对较弱,需要中心库的支持。在这种环境下,我们选择采取了集中式联机联合编目设计方案。

三、系统设计

I 系统软件选择

1995 年社科院文献信息中心购买了由丹诚公司研制开发的“丹诚图书馆集成系统”,它是利用 *Internet* 作为广域通讯环境的新型软件,后经不断升级,经过本地化和二次开发,系统功能更加完善。该系统具有以下特点:

1. Client/Server 模式

■支持多种硬件平台。*Datatrans-1000* 的 *Server* 产品通过支持 *UNIX* 和 *WindowsNT* 两种主流的服务器平台,支持分布在广域网上种类繁多的硬件平台,构成一个分布式图书馆业务环境,为书目信息的广泛共享提供了适应性极强的解决方案。

■提高了软件信息运行效率。进行大部分的事物处理工作,仅仅在必要时才访问服务器,提高了运行速度,节约了通讯开销。

■*Client* 软件是工作人员最频繁使用的工具,充分运用了 *windows* 平台的图形界面环境使业务工作更加方便和生动有趣。

■*Client* 端运行于 *Windows* 系列平台是日常工作的业务环境,*Server* 端运行于 *Windows NT* 或 *UNIX* 平台,为 *Client* 端提供数据库和目录服务。

2. 全面应用 MARK 格式

广泛采用 MARK 格式,为各种数据之间的共享和超文本链接带来极大的方便,由此提供了通俗、自然的检索界面。

3. 直接利用 Internet 实现联机数据共享

联机数据共享,就是利用网络环境,实时有效地共享各种数据资源,而不必借助物理存储介质的运送来缓慢地交换数据。系统以 TCP/IP 为基本通讯协议,保证了在 Internet 环境下有效地实现广域网通讯。通讯服务器结合数据库内核,可以将任何一个数据库设置成在 Internet 上提供全球联机编目的共享数据库,为用户开辟帐户、定义存取权限、计价收费,并保证数据的安全性。

4. 以 Windows 为操作系统平台,并有一个主要的服务器产品以 Window95/NT 为操作系统平台。系统还具有高集成度、超文本、事件驱动的界面风格。

总之,本系统采用 *Client/Server* 结构,服务器对数据集中管理,系统具有并发控制等功能,处理各客户机提交的作业请求,保证了数据的完整性和一致性。客户端负责用户界面管理,前端数据处理通过 TCP/IP 和 RPC 方式存取远端数据库数据。客户机基本配置:80386 以上 CPU,8M 以上内存,100M 以上的剩余硬盘空间,Microsoft Windows 3.2/95/98/2000 中文版。本联机联合编目系统服务器端为中小型机系统,基本配置为:80486 以上 CPU,16M 以上内存,Microsoft Windows NT3.51(中文版)以上版本,100M 以上的剩余硬盘空间。在具体选择上,考虑到社科院数据量滚动增长的具体情况,以能存储机读书目数据 40 万条为限,我们选择了 500 双主频,内存 256 可扩充,182 热插拔硬盘,8 插槽的服务器。联机编目服务器:主要实现主书目库各种 MARC 数据的查询、新增、修改、删除等操作,以及与之相关的索引文件和数据库文件操作。主服务器设有社科院书目总库、社科院书目预备库、社科院书目临时库。客户端联机联合编目软件:通过 TCP/IP 和

RPC 方式查询、调用、全屏幕编辑修改、删除远端数据库中的记录,并且具有本地处理的功能。可在本地建立小型的书目数据库,实现小型图书馆的书目查询、编目作业。系统提供数据批式上传和下载、打印卡片、数据联机查错、数据转入转出等多项实用功能。

II 联机联合编目作业系统流程和规范

1 系统结构

社科院书目总库:建立在互联网基础上的、庞大的、不断增长且经过筛选的高质量数据库。该库特点:质量较好,经过了查重操作,追加了馆藏字段。即每种书只有一条记录,每条记录反映了不同的馆藏地点。

社科院书目预备库:存储各所经过校验后的数据,只对各所三级帐户开放。

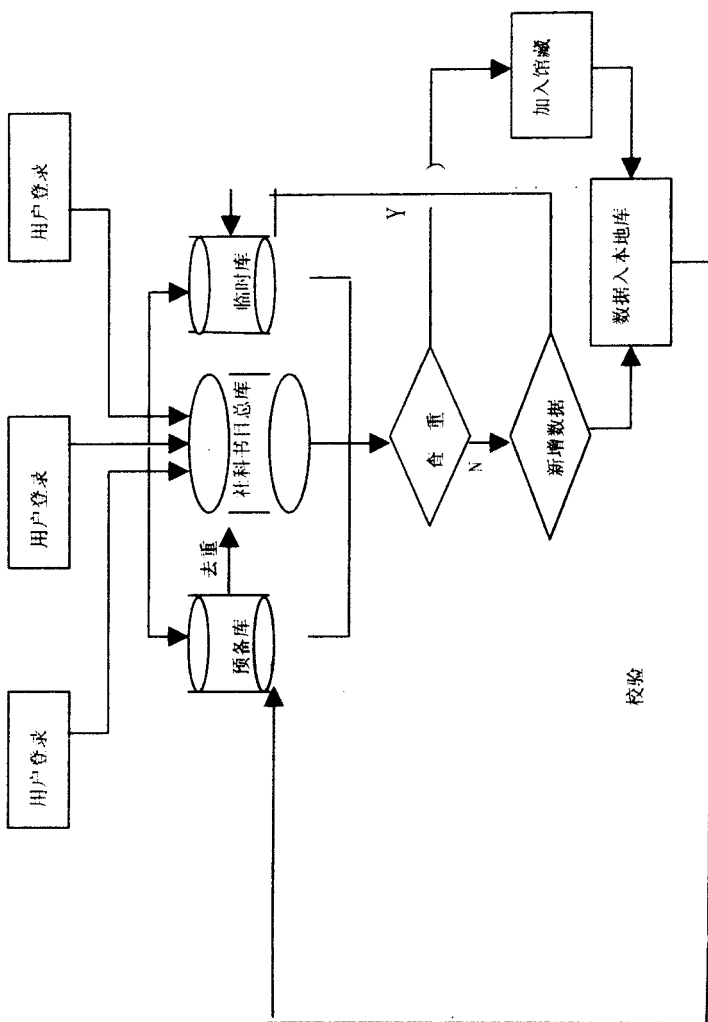
社科院书目临时库:为提供全院最实时的联机编目数据共享,参加联机编目的编目员对此库有写权限。编目员做好一条新记录后,保存在本地库的同时,应及时存至临时库,以便大家实时共享。

临时库的特点:动态、实时。

2 联机联合编目工作流程

一个典型的联机编目过程用流程图显示如下:

联机联合编目作业系统流程图



流程说明:

(1)用户登录: 用户启动联机编目客户端软件, 系统对用户的用户名、口令进行验证, 合法的用户方可进入联机编目状态, 否则只能在本地进行编目。

(2)查重作业: 系统提供多种检索途径查重, 并可按照用户的检索条件依次查询社科院书目总库、社科院预备库、社科院临时库。

(3)联机编目作业

● 如果重复, 调出该记录, 加入本馆的馆藏字段, 检查确认后存入本地书目数据库。经各所三级帐户校验后批存入预备库中, 通过社科院联机编目书目总库数据管理员进行查重、字段合并、字段追加等一系列操作, 将数据转入社科院书目总库。

● 如果不重复, 选择新增作业做详细的著录加工; 编目人员按照有关规范生产合乎质量的数据, 检查确认后存入本地书目数据库(即工作数据库), 同时将该数据实时存入社科院临时书目库。

(4)数据质量控制: 运用了质量等级排序的程序和采用机器与人工抽查的方式对数据库的质量进行控制, 如剔除重复数据、纠正格式和内容上的错误、增补不完整的数据等, 并定期提出质量评测报告和修正意见反馈给联机编目人员。

2. 联机联合编目所遵循的标准和规范

联机联合编目中心必须遵守现行的有关编目标准、通用的编目规范和本联机编目系统规定的有关规范。

(1)著录标准: 中国文献著录条例

(2)机读目录格式: CN-MARC

(3)主题词表: 参考中国分类主题词表或汉语主题词表

叙词标引规则: GB/T3860-95

(4)分类法: 中国图书馆图书分类法

(5)数据交换标准: ISO2709

(6)联机编目执行规范: 中国社会科学院联机联合编目手册

III 数据库的规划与维护

1. **数据库模式:** 设一个社科院联机联合编目总库、一个预备库和一个临时库。总库包括全院各馆的书目数据, 每个馆设一个分库。鉴于系统采用集中式设计, 结合社科院具体情况, 为了减少多用户同时访问主服务器时系统压力过大, 我们采用将联机联合编目总库放置于两个服务器上, 一个对文献中心内部提供访问和上、下载数据服务; 另一个对院外各所提供访问和上、下载数据服务。这样可以较好地缓解瓶颈问题。

2 数据库安全

数据库通过在服务器端设置帐户及口令, 保证数据库的安全。

3 数据备份:

3.1 库文件备份: 分别备份数据库的 5 个库文件, 以备系统崩溃时及时恢复。

3.2 文件备份: 定期将数据从库中转出, 以 ISO 文件形式备份。

3.3 双服务器: 系统支持双服务器数据库映象功能, 启动映象程序时系统会自动在指定的服务器跟踪备份。可在主服务器出现故障时, 及时顶替。在备份服务器与主服务器所有数据库设置相同的基础上, 将主服务器产生的日志文件, 实时提交给备份服务器, 备份服务器即可动态模拟主服务器上所有数据库的操作, 保持备份服务器与主服务器上数据一致。

4. **数据更新:** 初始建库时采用了查找与替换方式, 即对重复数据采取追加馆藏段的方法, 对不重复记录用追加入库的方法。采用三库, 大大保证了数据库中数据的唯一性, 同时又正确地反映了每本书的馆藏地。在以后的联机编目工作中采用实时联机作业和查找与替换相结合的方式。这是因为由于编目员

业务水平高低不同造成编目数据质量参差不齐,我们经过培训、考核对那些考试合格者给予实时联机编目的权限,而对于考试不合格者则采用对其原始编目数据存放于临时库中,供大家实时共享。对其套录的数据经本地数据管理员校验后转入预备库中,经社科院书目总库数据管理员进行查重、字段合并、字段追加等一系列操作后转入社科院书目总库。

四. 管理制度和人员培训

联机联合编目的系统方案和工作流程确定后,最困难的工作就是组织实施了。组织管理水平和管理措施的优劣决定着联机联合编目这项工程的成败。任何管理系统都是人力使用的系统,再由人力系统使用机械力系统,所以制订系统的发展目标和为确保达成此目标采取的控制措施是保证联机联合编目可持续发展的关键。

4. 1 联机联合编目工作的总目标:通过社科院文献信息中心和各个研究所图书馆的合作,建立社科院联机联合编目作业系统,实现社科院网上的书目数据共享。为用户提供准确、全面的书目查询服务和馆际互借服务。

4. 2 联机联合编目的组织管理

为了保证联机联合编目系统的正常运转,根据社科院机构设置和管理特点,我们采取两级的组织结构。

●第一层为管理和数据质量控制层。院文献信息中心于1998年10月成立了联机编目协调室,主要负责制订和阐明联机联合编目作业的发展目标;负责联机联合编目系统的正常运转,如数据的生产、管理维护和部门之间的协调等;负责制订数据质量要求和数据质量改进措施的全面执行;负责组织各方面专家研究和解决在数据格式和有关规则方面存在的问题;定期编制和印发全院联机编目工作简报,将课题进展情况和完成任务的情况及时通报