

●刘磊 郭世星

数字信息资源统一共享模式探讨*

摘 要 建立数字信息资源统一共享模式非常必要。数字信息资源统一共享模式,包括数字信息资源的生成标准模式、组织标准模式、存储标准模式、服务标准模式,以及建立统一的维护管理机构。每种模式都有其具体、深入的建设内容。图5。参考文献14。

关键词 数字信息资源 共享模式 网络环境 信息生成 信息组织 信息存储 信息服务

分类号 G253

ABSTRACT In this paper, the authors discuss the necessity of the unified sharing of digital information resources, and introduce some patterns for the sharing, such as generation standard pattern, organization standard pattern, storage standard pattern and service standard pattern for digital information resources. 5 figs. 14 refs.

KEY WORDS Digital information resources. Sharing patterns. Network environment. Information generation. Information organization. Information storage. Information services.

CLASS NUMBER G254

信息资源共建共享的模式一直是我国图书情报界关注的热门话题。在这方面,发达国家已有成功的经验,如 OCLC 模式、WLN 模式等;在技术标准研究上,则有 Dublin Core、XML 语言等方面的研究与应用。目前我国在这方面的实践和研究也已取得一些初步经验,如上海、广东和江苏等地区性数字信息资源共享模式,高校系统的 CALIS 模式等。然而从总体上看,迄今为止,我国在数字信息资源共享方面还没有一个理想的成功模式。

1 建立统一共享模式的必要性

1.1 数字信息资源利用存在的问题

近年来,以用户需求为驱动力的因特网在我国得到迅速发展,网络数字信息资源也日益增多。然而,由于诸多因素的制约,数字信息资源在其易用性、共享性、深加工方面的矛盾变得十分突出,其表现在:

- (1)格式各异,不能互通;
- (2)重复建设,资源浪费;
- (3)信息组织有序性差,利用困难;
- (4)数字化信息服务与传统信息服务分离等。

1.2 建立统一的信息资源共享模式的意义

(1)降低成本。由于拥有数字信息资源的机构各自为政,造成了大量的资源重复购置,加大了信息资源利用的成本。如果能建成统一的资源共享模式和标准服务体系,这种现象将得到很好的改善。

(2)整体性强。分散的信息资源对用户来说是极其不方便的,利用效率也不高。基于网络环境下的数字资源共享模式,虽然同样由分散的部门所组成,但对用户来说却感觉不到这种组成的痕迹。

(3)效用性高。数字信息资源的共享使人类获取信息发生了根本性的变化,信息资源的数字化和网络化使人们存取文献的速度和范围得到了极大的提高和扩大^[1]。

2 统一共享模式的设想

网络环境下数字信息资源共享模式的建立应将处于不同地理位置、属于不同专业领域、处在不断更新状态中的各种数字信息资源用网络链接起来,以便于对网络环境下信息资源进行存取和利用。本文探讨的网络数字信息资源共享模式是从共享过程的五个组成要素入手来展开的,这五个要素包括数字信息生成、数字信息组织、数字信息存储、数字信息服务和整体过程维护管理。如图1所示。

* 本文系南京农业大学人文社科基金及南京农业大学信息科技学院课题资助项目的研究成果。

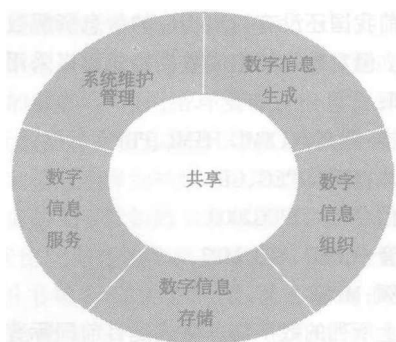


图1 数字信息资源共享过程要素

数字信息的生成是把各种信息资源通过不同的标准或协议转换成网络环境下计算机统一识别的数字信息资源,即将分散于不同载体、不同地理位置的

信息资源以数字化形式存贮、以网络化方式交互传递,实现资源共享,从而最大限度地满足人们获取信息、利用知识的需要。数字信息资源的组织即通过分类、主题、分类主题一体化等组织方式为数字信息资源提供有序化的结构,使之形成一个有机化的整体。数字信息资源在有序化的基础上必须有足够的存储空间和简单快速的存储算法,从而有利于信息资源的充分快速的利用。所有的资源所进行的一系列的处理最终的目的都是能够为用户提供满意的服务。数字信息资源共享作为一个完整过程就必然要进行维护管理。对以上五个模块有个初步了解以后,图2介绍了五个模块之间相互配合的数字信息资源共享过程。

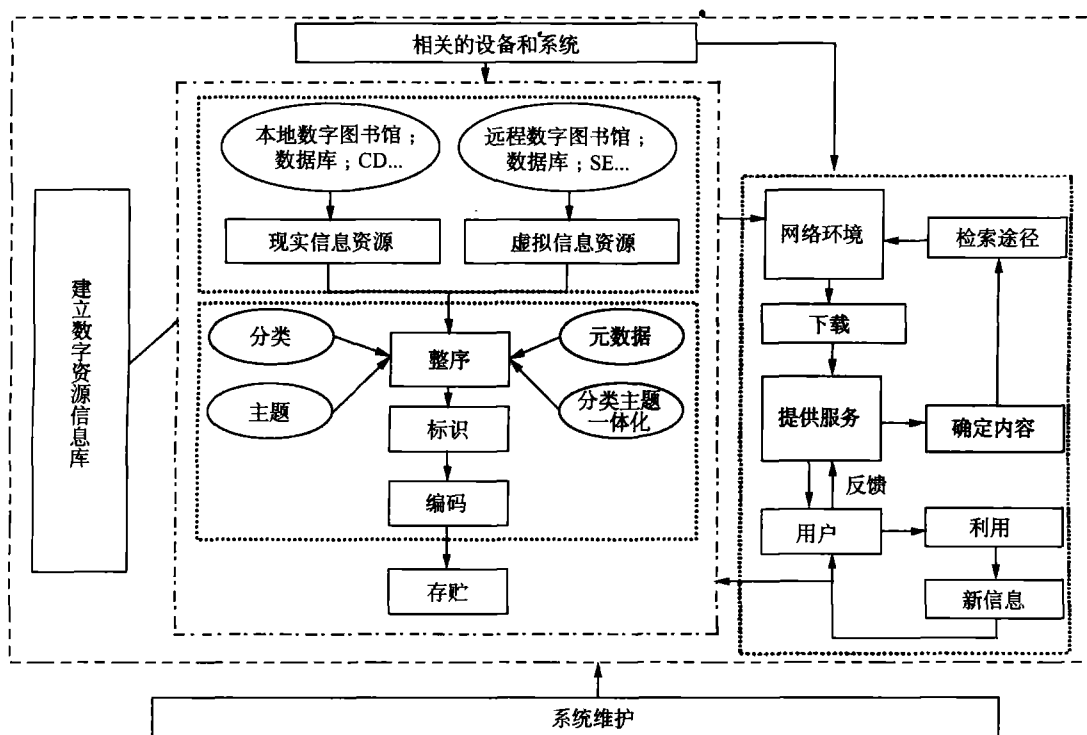


图2 网络环境下数字信息资源共享过程

由图2可以看出,数字信息资源的生成可分为现实资源(放置于本地的数字信息资源)和虚拟资源(通过网络才能获得的数字信息资源)^[2]。两部分的获得途径有:传统文献资源进行数字化加工处理;补充建设新型媒体电子出版物,如CD-ROM、VCD、DVD等光盘文献、音像制品、教学软件、电子书刊等;搜集、筛选、下载镜像网上信息;自主开发建设学术科

研成果数字化文献资源;建立导航网站,链接网上专业网站、搜索引擎、数据库服务中心、数字图书馆等建立虚拟网上资源系统^[3]。然后通过信息资源的整序、标识、编码等过程完成信息组织。整序的方式可以根据被组织信息资源本身特点选择采用哪种方法来进行组织,如元数据方式、分类法、主题法、分类主题一体化等组织方法。这样我们就有望建成一个统

一的数字资源数据库,并存储于网络中,供用户利用,且在共享的过程中进行系统的维护管理。

以上是给出的网络环境下数字信息资源共享模式的一个总体雏形。

2.1 数字信息生成的标准模式

国家有关部门应组织信息产业界、图书情报界和国内软件开发商参与讨论制定电子书刊标准、各种元数据标准、多媒体信息等标准,尽快实现数字化信息资源的标准化建设,以便在统一的标准和协议下,开展分布式海量信息资源建设与检索应用。

信息资源数字化建设应包含两方面含义:其一是对原有传统文献资源进行数字化加工转化,其二是开发建设新的数字化的信息资源。要对二者进行科学的整合、重组、分类、组织,形成新的数字化资源体系。

数字化信息的生成技术包括键盘输入技术及非键盘输入技术。非键盘输入技术又包括手写识别技术,印刷文稿扫描与识别技术,语音识别技术等^[4]。不论采用的是哪种生成技术,我们要求生成的数字化信息能够是采用国际上普遍采用并相对开放的标

准。目前我国还没有提出相应的信息资源数字化制作标准。但对最终的生成数据格式可以采用国际标准^[5],如:

文本:纯文本、XML、HTML、PDF

图片:TIF、JPEG、GIF

图像:JPEG、JPEG2000

声音:MPEG、AC3、MP3

视频:MPEG - II、IV

以上所列的数字信息格式是目前国际普遍采用的几种数据格式。在没有更好更新的统一数据格式之前,我们要想达到信息资源的最大共享就应尽量采用这些国际通用的标准。

2.2 数字信息资源组织标准模式

信息组织是将处于无序状态的特定信息,按照一定的标准和方法,使其成为有序状态的过程。其目的是将无序信息变为有序信息,便于人们利用信息和有效地传递信息。网络环境下数字信息资源组织方法可以从三个方面——语法、语义、语用入手去分析(如图3所示)。

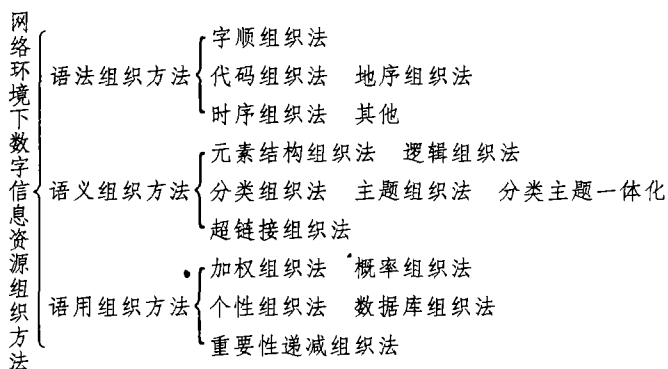


图3 网络环境下数字信息资源组织方法

语法组织方法是语义组织方法的一种外在表现形式,语用组织方法是从用户检索角度去分析信息资源的组织方式,语义组织方法是对网络环境下数字信息资源内容的揭示。以下主要分析语义方法中分类法、主题法和元数据方法。

(1)分类主题一体化在网络环境下数字信息资源组织上的应用。

在网络环境下,单纯的使用分类法或是主题法组织信息都有各自的优势和不足,都满足不了网上信息查询的需求。理想的模式是:“分类——主题相结合的信息组织方法,即按类别先类分信息,再在类

别中以主题方式加以组织,从而为不同层次和需求的用户提供分类的、主题的、分类主题一体化的信息检索途径。分类主题一体化包括三个方面的内容:一是建立一个结构简明的知识分类体系。通过对信息的系统分类,实现对信息的宏观控制,当信息分别组织到各个类中之后,其含义就受到了类目的限定,为过滤不需要的信息创造了条件。二是建立一个智能化的受控主题词表,实施对作者语言与用户语言的控制和转换。三是建立分类体系和主题词表的联系系统。即将标引语言纳入分类体系^[6]。这样会接近用户的要求,大大提高检准率。

(2)元数据在信息组织的应用。

元数据可以理解为“关于数据的数据或描述其他数据的数据”。在网络环境下的信息资源的组织方面,元数据具有信息描述、信息定位、语义搜索、信息评估和信息选择的功能。元数据与网络环境下的信息资源的关系就如同书目数据与图书之间的关系一样,能比较完整地反映出信息对象的全貌^[7]。因此,在图书馆和信息技术界,正在利用 DC (Dublin Core)元数据格式对网络环境下的信息资源进行编目。其核心是如何用一个简单的、灵活的、标准化的元数据记录来描述种类繁多的网络环境下的信息,使非图书馆专业人员也能了解元素含义并熟练使用其描述格式。DC 元数据集是元数据格式发展中比较成熟和成功的一种。DC 元数据标准的创建,是通过网络环境下的信息资源的提供者对资源属性信息的描述并对其内容进行编目来帮助人们尽快地在网上发现所需要有效信息资源。因此,DC 元数据资源描述方案又称为网上资源的编目。由于 DC 元数据已被结构化,并支持字段查询,为从根本上解决网络环境下的信息资源的编目问题提供了切实有效的方法^[8]。

2.3 数字信息资源存储标准模式

(1)网络环境下数字信息资源的物理存储标准模式。

目前,以服务器为中心的信息系统正逐步向以网络、数字信息为中心转移。传统的以服务器为中心的直存储技术 DAS 正在向以 NAS 和 SAN 为代表的网络存储技术发展。用户可以在网络上方便地存取数据,利用客户端浏览器进行访问和管理^[9]。

在国内,NAS 网络附加存储(network attached storage,NAS)和 SAN 存储区域网(storage area network,SAN)是比较新的概念,作为两种主流存储技术的出现,从根本上改变已有的存储结构和存储管理方式,为避免大量数据传输造成的网络拥塞,实现数据集中管理,提供了有效的解决方法。两者有着各自的优势,但也各有弊端存在。NAS 与 SAN 技术拥有不同的市场,未来的信息机构并不是只需要某种单一的技术。将两者融合,互为补充,扬长避短,成为目前的一种发展趋势。从信息存储操作单位看,NAS 是以文件存取,功能偏重文件共享,SAN 是以块操作,主要功能是高速存取,融合的关键是能够将两者统一起来管理,为用户提供一个统一的存储资源池。未来的目标是将 SAN 连接到 NAS 上,使网络存储共存

于一个存储网络体系中^[10]。两者的结合在系统的可用性、资源共享、存储合并以及服务模式等方面将有很大的提高。

(2)网络环境下数字信息资源的逻辑存储标准模式。

文本、图像、音频、视频等信息资源数字化处理后,以多媒体数据库(MDB)的逻辑结构实现存储,并用多媒体数据库管理系统(MDBMS)实现管理。在网络环境下根据存储数字信息的特点本文采用多媒体数据库系统实现技术中的三种:扩充关系数据库技术,面向对象的多媒体数据库技术及超文本或超媒体数据库技术来实现数字信息资源的逻辑存储标准模式^[11]。

①扩充关系数据库技术。传统的关系型数据库系统(RDBMS)只能描述字符和数字等符号信息,无法描述图像、动画、视频、声音等多媒体信息。扩充关系数据库的方法就是在 RDBMS 的基础上引入新的数据类型以描述多媒体信息。可见,扩充关系数据库技术仍然主要用于描述文本信息。

②面向对象的多媒体数据库技术。面向对象的多媒体数据库 OOMDB 从多媒体的数据模型入手,采用面向对象中对象、方法、属性、对象类的层次结构和继承特点来描述多媒体的数据模型。

③超文本或超媒体数据库技术。超文本采用非线性技术组织和表示块状信息。结点和链接超文本的两个核心。利用超文本实现 MDBMS 的相关技术有超媒体问题,多媒体检索问题,版本控制问题和标准化问题。超文本或超媒体数据库技术主要是解决如何用结点和链接来组织和表达多媒体及其相互关系以及如何实现超媒体的同步和协调等。

2.4 数字信息资源服务标准模式

网络环境下数字信息资源的服务方式可分为基本服务和扩充服务。基本服务主要是指浏览检索服务。扩充服务是指由某些专用的应用软件或用户接口提供的服务方式,它包括个性化服务、集成化服务、定题跟踪服务、信息增值服务、参考咨询服务等^[12]。网络环境下数字信息资源服务标准设想的基本模式如图 4。

由图 4 可知,用户通过统一的检索界面发出不同的服务请求,经过系统的分析发送给服务器不同的检索策略,服务器根据检索策略到数字信息资源中找得命中的信息,返回给用户。用户在利用命中信息之后会产生新的信息,通过反馈机制收入数字信

息资源库^[13]。具体的服务内容如下:

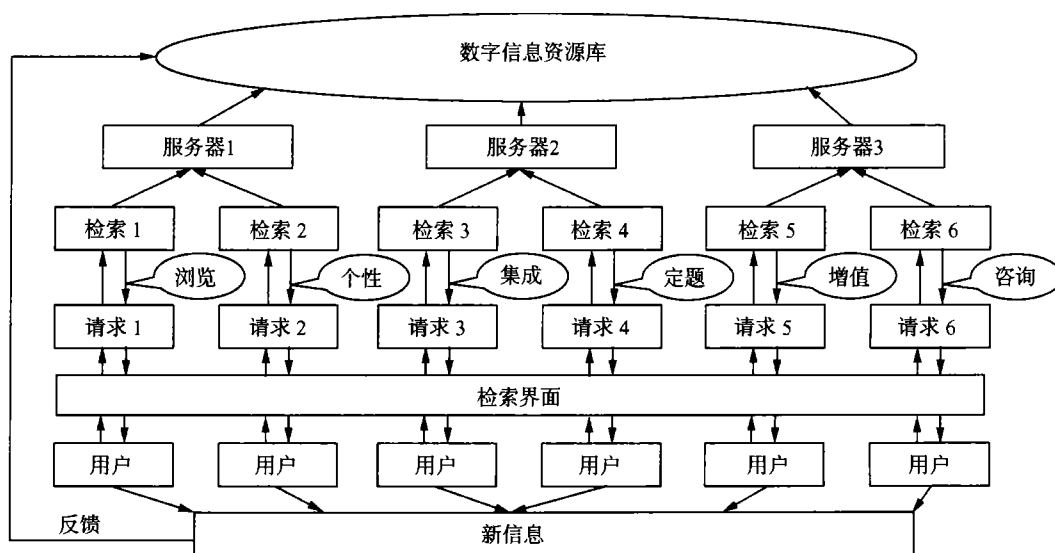


图4 数字信息资源服务标准模式

(1)浏览检索服务。建设方便实用的信息机构主页,面向各类网络用户提供开放式服务。利用主页提供的丰富类目,可浏览、查找、下载相关信息;利用系统提供的全文检索系统,可对各类信息机构所有的数据库进行全文检索和下载。

(2)个性化服务。根据用户提出的请求,工作人员利用网上丰富的专题性、学科性信息资源,采用信息推送技术直接推送到用户的桌面,实现个性化服务。

(3)集成化服务。针对用户的需求,将分布各种载体上的信息进行集成,使信息有序化、浓缩化、精细化、专题化、知识化,重组成新的信息资源体系,利用网络或光盘方式提供给用户。

(4) 定题跟踪服务。以选定的重点研究课题为服务目标,例如重点学科建设、重点科研课题、工程技术项目论证等课题,从课题立项到成果鉴定,在网络环境下进行全程跟踪服务,提供针对性强、时效性高的信息检索和信息分析服务。

(5)信息增值服务。根据用户提出的请求,搜集本单位和网络上的信息资源,经过综合分析研究和加工处理后,形成信息产品,包括光盘产品、专题信息研究分析报告、定题信息资源库等,实行有偿服务。

(6)参考咨询服务。在网上建立虚拟参考咨询

组,及时解答用户提出的有关网络利用、网络资源分布、信息服务、信息检索等方面的提问;开辟参考咨询服务专栏,将最常见的问题和解答编制成文献存入主页,使用户自由访问。

2.5 建立统一的维护管理机构

可以考虑以下统一的系统维护管理模式(图5)。

如图 5 所示,应该在国务院直接领导下以信息产业部和有关部门为基础成立国家信息化委员会和相应的监督机构,统一管理维护我国信息资源建设(包括数字信息资源)^[14]。各省、市、自治区建立地区管理机构,负责管理各地区下属的图书馆系统、信息系统、档案系统和各类商业信息组织(如 ICP、数据库开发商)的信息资源建设。为充分合理地配置利用分散在全国各地区和各系统的信息资源,要建立全国和地区统一的资源配置中心,统筹规划各地区各系统的信息资源的配置与开发。根据发达国家的经验,我国数字信息网络共同体应该建立在网络联合目录的基础上。要在政府部门的帮助下,首先开发个别机构之间的互借信息系统和物流系统,解决其自动联络和费用结算问题,然后把系统的范围扩大到整个地区乃至全国。建立数字信息资源共享机制的关键在于能否充分发挥政策、法规和市场调节的作用,形成合理的具有互惠互利经济结算机制的数字信息资源利用的共同体。

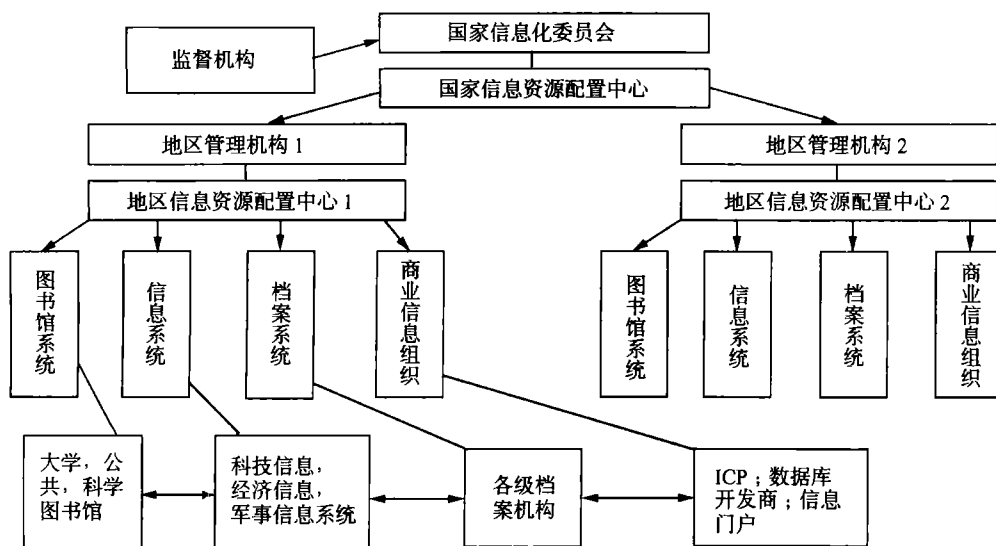


图5 统一的系统维护管理模式

参考文献

- 1 吴慰慈, 张久珍. 网络环境下的信息资源管理的标准化体系研究. 情报科学, 2001(1)
- 2 金萍, 王彬. 浅谈信息网络时代信息资源的共享. 东北水利水电, 2001(2)
- 3 吴志荣. 数字图书馆: 从理念走向现实. 上海: 学林出版社, 2000: 36 ~ 172
- 4 刘阳, 丁银燕. 论图书馆信息资源数字化建设. 图书馆工作与研究, 2002(2)
- 5 黄波. 信息资源数字化建设与应用技术. 现代图书情报技术, 2003(1)
- 6 储节旺, 郭春霞. 论网络信息组织. 情报理论与实践, 2000(4)
- 7 付利庆. 网络信息资源的组织和揭示初探. 见: 《图书馆杂志》编辑部, 上海市图书馆学会. 网络资源的组织与管理. 北京: 北京图书馆出版社, 2002: 9 ~ 16
- 8 郝亚玲. Metadata: 数字图书馆信息资源的整序. 图书馆理论与实践, 2002(5)
- 9 刘颖, 李静霞. 网络信息组织与发现的有力工具: Dublin Core. 图书馆建设, 2002(1)
- 10 史田华等. 信息资源组织与存储. 南京: 东南大学出版社, 2003: 267 ~ 292
- 11 张春雅. 网络环境下的信息存储主流技术及发展趋势. 中国信息导报, 2002(12)
- 12 郑建民. 网络信息资源的组织方式与开发利用. 江西图书馆学刊, 2003(4)
- 13 郭振按, 牟建国. 网络环境下信息资源重组与服务模式

研究. 河北科技图苑, 2003(6)

- 14 刘磊. 网络环境下我国社科信息事业发展的区域协调问题. 中国图书馆学报, 2002(6)

刘磊 博士, 南京农业大学信息管理学系教授, 硕士生导师. 通信地址: 江苏南京. 邮编 210095.

郭世新 南京农业大学信息管理学系硕士研究生. 通信地址同上.

(来稿时间: 2004-07-13)

敬告读者

由于信息传递失误, 致使蒋永福先生《知识权利与图书馆制度——制度图书馆学研究》一文在本刊(2005年第1期)和《图书馆建设》(2005年第1期)上同时刊出, 并非作者一稿两投. 特向广大读者致歉。

《中国图书馆学报》编辑部

2005.1.25