

●马大川 杨红平

## 信息资源的集成整合研究\*

**摘 要** 信息资源集成的目的,是对若干个信息制作、管理单位中分散存储和管理的各类信息资源,利用计算机技术、数据通信技术、网络技术和数据库技术进行统一管理,实现信息资源的数据共享。它要遵循实用性原则、用户至上原则、系统性原则、标准化原则等。图3。参考文献15。

**关键词** 信息资源 资源集成 资源整合 信息资源建设

**分类号** G253

**ABSTRACT** The objective of the integration of information resources is to manage separately stored and managed various information resources in a single system by using computer technology, communications technology and network technology, to realize the sharing of information resources. We should adhere to the principles of practicality, "user is the god", systematicness and standardization. 3 figs. 15 refs.

**KEY WORDS** Information resources. Integration of resources. Development of information resources.

**CLASS NUMBER** G253

### 1 问题的提出

随着信息化浪潮在全世界的兴起,信息资源的重要性越来越明显。在我国,信息资源建设与应用技术也有一定的发展,但由于缺乏完整的理论体系、统一的思想认识、相应的标准和规范、科学的规划和组织,目前存在以下问题:

(1)信息资源建设缺乏整体规划。目前的一些信息资源加工项目,建设思路仍未摆脱传统的窠臼,信息资源建设偏重于文艺、外语、计算机、经济等方面的书刊报资料。

(2)学科内容不均衡,信息资源建设缺乏连续性,资源分布不成体系并缺乏科学性,整个资源体系杂乱无章,没有条理。信息应该从广度和深度两个方面进行高度集成化,否则不足以形成资源,不足以形成市场。

(3)资源再生性与二次利用率差。资源的再生性和二次利用率是考察信息资源建设成败与否的一个关键指标。片面追求数量和速度,不考虑资源的再生性与二次利用,不利于信息资源的可持续发展 and 高效共享。

(4)缺乏信息资源使用的增殖意识。

(5)信息资源地域分布不均衡,存在大量重置和浪费现象;信息资源基础设施软、硬件平台不统一,造成资源不能共享。

(6)资源关联度差,信息资源库、应用系统异构程度高,造成大量“信息孤岛”<sup>[1]</sup>。

由此不难看出,传统的建设模式带来的结果是:一方面解决问题,另一方面制造信息孤岛;一方面各级管理部门苦于基础数据库没有信息、没有信息渠道和更新渠道,而另一方面大量已掌握的第一手、真实的数据又没有被充分有效地利用。这种状况不仅使人们投资建设的许多信息采集、通信和应用系统没有充分发挥作用,也制约了应用的进一步发展。

资源集成整合是突破现状的惟一途径。怎样才能突破目前的局面,实现信息资源建设的可持续发展呢?若将现有系统推倒重来,建设综合的新系统,是不可能也是不现实的。现阶段要花钱少、见效快地大幅度提高行业信息化和信息资源建设质量和信息技术应用水平的惟一途径就是整合、集成现有资源。

### 2 信息资源集成整合的原则、特点和意义

\* 本文系教育部人文社会科学重点研究基地重大项目(02JAZJD870003)的研究论文之一。

集成是一种理念,它有多种称谓,如:“集成”、“整合”、“一体”、“综合”等。所谓集成,是指某一系统或某一系统的核心把若干部分、要素联接在一起,使之成为一个统一整体的过程。从管理的角度来说,集成是一种创造性的融合过程,只有当构成一系统的要素经过主动的优化、选择搭配、相互之间以最合理的结构形式结合在一起,形成一个有适宜要素组成的、优势互补的有机体,才能被称为集成。信息资源的集成不是信息资源的堆积或信息载体的物理堆积,其本质是一种竞争性的互补关系,即各种要素通过竞争冲突,不断寻找、选择自身的最优功能点,在此基础上进行互补匹配<sup>[2~3]</sup>。信息资源集成的目的,是对若干个信息制作、管理单位中分散存储和管理的各类信息资源,利用计算机技术、数据通信技术、网络技术和数据库技术进行统一管理,实现信息资源的数据共享。在集成过程中,还必须保证共享资源库中数据冗余最小。

## 2.1 信息资源集成整合的原则

在数字化的条件下,由于网络传递的便利性,使得信息资源集成建设的特色性原则显得更为重要,除了要遵循一般信息资源的建设原则诸如实用性原则、用户至上原则、系统性原则、新颖性原则外,还应具备一些适应数字化网络化环境的特定原则。它们是:

(1)整体性原则。是指要保持信息资源对象学科的完整性。整合后的资源系统应涵盖各子系统内部功能,反映数据对象间的内在关系。

(2)标准化原则。信息资源集成的标准化主要包括数据格式的标准化、描述语言的标准化、标引语言的标准化,还有通讯协议的标准化、安全保障技术的标准化,以及数据管理软件、硬件的标准化,以保证信息资源的共建共享。

(3)优化性原则。是指运用一定的技术手段和方法,使信息资源得到合理组合,取得最好的组织结构和组织功能。

(4)安全性原则。计算机病毒、黑客、软件炸弹、信息垃圾、存储设备故障等方面的问题给信息资源集成带来了极大的安全威胁。在信息资源的集成建设中,应采取必要的安全保障措施来保证信息资源的安全<sup>[4]</sup>。

## 2.2 信息资源集成整合的特点

信息资源集成建设采取科学、合理的整体规划,资源能连续更新。它将各种数据库、光盘资源、网上信息资源、电子格式文献、印刷型文献以及视觉类信

息等不同种类文献真正融合在一起,便于读者检索使用。

采用大型数据库、新型检索技术、智能代理、全文搜索引擎等重新整合信息,提高显现知识含量,采用自然语言理解技术提高文本存储和检索结构的精度,采用网络技术、COM 组件技术等保证系统的开放性扩充,实现信息表现、传输、加工的智能化、自动化和规模化,都是信息资源集成的特点。概括而言:

(1)整合后的资源来自不同学科的数据库,因而知识覆盖面较广,能够提供系统的学科知识。

(2)整合后数据对象间具有统一性和有机关联性,体现了资源的系统性。这种关联性反映了知识间的内在联系。

(3)整合后的信息资源具有相同的组织结构和组织功能,提供统一的用户界面和共同的检索方法。

(4)整合后的信息资源在结构功能上的统一性便于数据的更新和扩充<sup>[5]</sup>。

## 2.3 信息资源集成整合的意义

(1)解决信息资源分散性与读者信息需求综合性的矛盾。当今,科学和技术向高度专门化和综合化发展。这种高度专门化与综合化的趋势,使读者的信息需求呈现综合化的特点。读者信息需求的综合化表现在:一是信息需求内容的综合化;二是信息需求的全方位发展。这两个方面的要求集中体现为读者愈来愈需要得到内容全面、类型完整的信息。但是信息资源分布的规律是:“大量该学科的文献相对地集中在一定数量的杂志上,而剩余部分的文献则依次分散在其他大量的相关杂志上”。特别是现有的网络检索工具,搜索范围覆盖面广,专指性差,人们在检索学术信息时,往往从专门学科入手,而网络检索工具却不以专业划分,在查询某一特定学科网络信息时,一方面没有专门的网络检索工具,另一方面又是任何一个网络检索工具都可以用,致使检索的完整性可靠性都较差,难以实现对网络资源的有效过滤和筛选。整合的目的就是要使信息用户快速获取知识<sup>[6]</sup>。

(2)提高了信息资源的利用率,实现了信息的增殖服务。丰富的数字化资源,高速的网络传输,使用户可以在世界上的任何地方任何时间查询多个分布式信息资源,大大缩短了信息传递时间,拉近了信息提供者和使用者的距离,加快了信息交流与反馈的速度,提高了信息资源的利用率,实现了信息的增殖

服务,创造了规模化经济效益和社会收益。

(3)实现最大程度的资源共享。真正从信息用户出发,突破信息资源地域和时间上的限制,建立数字信息资源的协调化、集成化体系,解决有限的信息收藏和无限的用户信息需求之间的矛盾,最终实现资源共享。

### 3 信息资源集成整合的实现

#### 3.1 应用系统的集成

信息资源服务结构的系统构建包括软件和硬件。软件包括支持硬件运行的系统软件以及工具性的和应用性软件。如浏览器软件、多媒体应用软件、搜索引擎软件等。硬件包括网络通讯设备、PC 及各种接入输出设备等。信息资源系统随计算机硬件的发展而发展,由最初的集中式系统不断加入新的计算机和 workstation,进而发展为网络系统。这样,就必然同时存在着大型机、PC 和 SUN workstation,运行着 Win3x、Win9x、WinNT 以及 UNIX 等操作系统和 Sybase、SQL Server 等各种数据库。利用网络互联,

就可能共存着以太网、令牌环网、PSTN、FDDI 网、ATM 网等,以及应用于其上的各种通讯协议,包括 TCP/IP、IPX/SPX、NetBEUI 等。应用系统的集成就是要在异构计算机、异构网络环境下实现交互操作,方便信息的共享和访问。

当前的主流集成技术大多采用面向对象的多层客户机/服务器(C/S)模型。该模型将分布在网络上的资源都按对象的概念组织,每个对象都有明确的访问接口。基于 C/S 模型的 CORBA 技术很好地解决了系统异构的问题。

对象管理组织(Object Management Group, OMG)于 1991 年提出的通用对象请求代理体系结构(Common Object Request Broker Architecture, CORBA)的技术规范,就是为了解决软硬件协同工作而提出的很好的方案。CORBA 借助于客户/服务器计算模式和面向对象技术,为分布异构环境下各类应用系统的集成,实现应用系统之间的信息互访、知识共享和协同工作提供了良好的可遵循的规范技术标准。其结构如图 1 所示<sup>[7~10]</sup>。

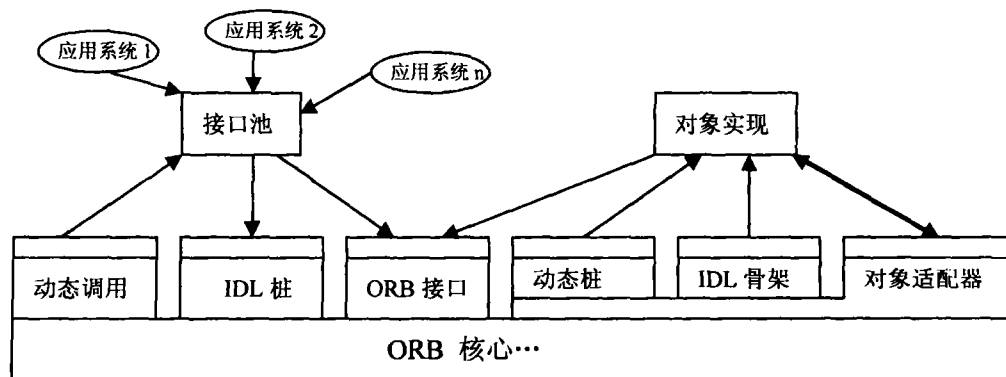


图 1 CORBA 结构示意图

CORBA 的核心是对象请求中介(ORB),它定义了异构环境下对象透明地发送请求和接收响应的机制,客户机不需要知道请求对象的位置、使用的编程语言、通讯协议等,它所关心的是找到某个对象实现这个请求,然后传递参数和方法,并将结果返回给服务器对象。接口定义语言(IDL)独立于编程语言,用于描述对象和操作。IDL 提供了对成员系统的封装和成员系统之间的隔离,任何成员系统作为一个对象,可通过 IDL 对其接口参数进行定义和说明,就能连接到 ORB 上,为其他系统提供服务或向其他系统提出请求,方便地实现了跨平台的通讯和客户机/服务器对象间透明交互的能力。CORBA 提供了软件

总线(Software Bus)机制,支持组件对象的即插即用,具有跨网络的互操作能力与可移植性,可以并行地运行在多个系统平台上,能够实现多目标系统间的无缝连接。

目前各计算机厂家纷纷提出了符合 CORBA 标准的 ORB 产品,如 DEC 公司的 ObjectBroker、HP 公司的 ORB.Plus、IBM 公司的 DSOM、Iona 公司的 Orbix 等。

#### 3.2 资源数据库的整合

由于互联网自由、开放的特点,导致网络信息资源的混乱、无序。由于原创网络信息资源的贫乏和分布结构的不平衡,类型的异构,网络信息资源存在

大量的重复、过时、凌乱的问题。对信息资源数据库进行整合,就是要处理各种不同结构的数据,即结构化、半结构化和非结构化的数据类型。

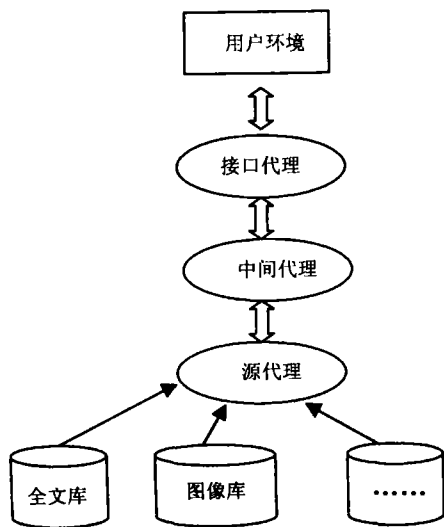


图2 基于代理的集成结构

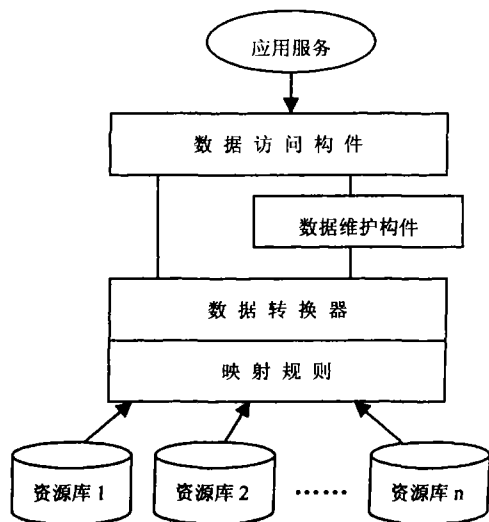


图3 异构数据库的集成

解决上述数据库资源类型异构的问题,一种方法是采用基于中间件的代理技术,如图2所示。中间代理包括接受用户查询请求,定位相关信息源发送请求等。当传送一条查询请求时,中间代理把它发送给相邻代理,如果该代理不能处理,转发给下一代理,直到能够处理<sup>[11~12]</sup>。另一种方法是建立实际资源数据库与资源概念数据库的映射规则。集成体系结构如图3所示<sup>[13]</sup>。所谓资源概念数据库就是资源

数据库的视图;数据库转换器通过映射规则,读取相应的映射描述文件,转换后反映到相应的应用系统数据库中。这样,通过对数字资源数据的抽象建立实际资源数据库到资源概念数据库的映射规则,提供一个统一的数据视图,具体资源数据库的结构和表示方式对用户透明,屏蔽了不同数据表示差异对应用的影响。对用户来说,得到的是一个整体的数字资源库,各种查询和访问都可以针对这个整体资源库进行,从而实现了不同数据库的集成。

### 3.3 建设人才的整合

信息资源集成是科技含量很高的知识创新工程,技术性很强,必须组建一支高素质的,能支撑整个系统运行的,具有深入探索研究能力的专业技术队伍。需要有多方面知识结构的人才,特别是在计算机、自动化和网络方面具有坚实基础的专家,和在数字资源开发、组织和提供利用方面具有丰富经验的管理人员。因此建议采取以下策略:(1)树立以人为本的管理观念,把对人的能力的培养和积极性的发挥放在头等重要的位置。(2)改变组织结构,从传统的金字塔型组织结构改造成网络时代更能体现效率的扁平化的组织结构。(3)加大培训力度,树立长远的教育战略思想,形成一套科学的教育培训计划。(4)完善激励机制,包括人才配置机制、付酬机制、培养机制、奖励机制等<sup>[14]</sup>。

信息资源整合程度的高低不单单取决于人们的技术高低,更重要的是信息服务人员有没有整合的意识。只有培养和造就一批高素质、高水准的信息服务人员,才能随时为用户提供个性化、及时性的帮助服务,培养用户获取信息与知识的能力,帮助用户了解掌握网上资源的组织结构与方式,熟悉网络资源库的分布状况,发挥网络信息导航员的作用;或接受用户委托,代为检索加工信息,为用户提供智能化的成果型的知识产品;或为用户开展不同层次的培训,提高用户的信息素质,发展信息用户,促进信息资源的利用。

### 3.4 信息资源的安全管理

信息资源集成后既要保证信息的安全和保障信息资源拥有者的权益,又能使用户方便地完成数据检索和传输。这就要从法律道德、基础设施、电磁保护、网络结构、网络设备、软件系统、网络管理、密码体制、密钥保障等各方面入手,构造基于环境安全、网络安全和密码安全3个层次的完善的信息安全体系。

应当指出,以上这些层次的整合并不是各自孤立的,而是一项一体化工程。整合的最终结果应是为个人信息消费者、社会各级政府部门、企事业单位、社会服务机构等提供一个信息交流、内部协作、强化规范管理、提升原有资源价值的平台,以实现信息资源的高度共享和增殖服务。

#### 4 信息资源集成整合的标准化和规范化

信息资源集成的标准化和规范化需要全国范围内很多图书情报档案机构、各种信息中心和文化设施等众多部门和单位共同参与;它所涉及的信息和知识囊括了所有学科,数量极其巨大,类型特别繁多,而且包括了文字、表格、图像、音频等多种多媒体的数字化表达和无缝连接,组织极其复杂。如何将众多的力量协调组织起来,实现网络的互联互通、资源的共建共享、管理的井然有序,从技术管理的角度考虑,关键就在于标准化。标准化和规范化是信息资源集成与共享的一个十分突出的问题,并成为实现信息资源集成与共享的前提和根本保证。应当在充分调研的基础上,逐步推出比较完善的标准,如信息资源的储备、描述与标识、检索查询、交换和使用的标准与规范等,并使之成为标准体系的重要基础。架构信息资源集成的标准体系,可从以下几个方面进行考虑:(1)数字化信息采集标准。(2)数字化信息组织与存储标准。如文本信息的表示和存储、多媒体信息的存储、数字化信息的著录分类和标引、多媒体信息的压缩、网络环境信息资源标识、元数据标准等。(3)信息检索标准。如全文数据库检索、多媒体信息检索、异构系统的互操作标准等。(4)网络与网络资源标准。如传输控制及互联协议、信息资源网站评价、网络信息资源组织标准等。(5)信息的权限管理和安全标准。如加密、水印技术、指征鉴别等。(6)其他标准。如信息文献工作应用软件评价及评价指标体系、文献信息系统质量管理和质量认证体系等<sup>[15]</sup>。

#### 5 结束语

建立信息资源集成整合机制,这首先是由资源本身的多元化决定的,现有文献信息资源除了传统的印刷型载体外,还有电子资源和网络资源。其次是由读者的需求决定的,随着网络技术的发展,信息用户总是要求在一个地方即可查到所需要的各种类

型的文献信息,这就要求文献信息管理由传统的“多站式”向“一站式”发展。只有整合信息资源,为读者提供的服务才会更主动更方便。

通过信息资源的集成整合,可以最大限度地实现资源共享,尽可能地减少中间环节,缩短信息传递与反馈的速度,提高使用效率;从全社会来说,能够减少信息处理的费用,降低信息与信息服务的成本,也有利于充分发挥信息资源的整体效益和规模效益。

#### 参考文献

- 1 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)
- 2 李宝山, 刘志伟. 集成管理——高科技时代的管理创新. 北京: 中国人民大学出版社, 1998
- 3 霍国庆, 杨英. 企业信息资源的集成管理. 情报学报, 2001(1)
- 4 顾英, 席建平. 数字化信息资源建设若干问题的探讨. 现代情报, 2003(7)
- 5 马文峰. 数字资源整合研究. 中国图书馆学报, 2002(4)
- 6 冯花朴. 高校图书馆数字化信息资源的开发. 情报科学, 2003
- 7 刘成山, 赵捧未, 窦永香. 系统集成技术在数字图书馆中的应用. 情报理论与实践, 2003(2)
- 8 Vinoski S. CORBA: Integrating Diverse Application within Distributed Heterogeneous Environments. IEEE Communications Magazine, 1997(2)
- 9 隋秀凛. 基于 CORBA 的制造企业信息集成系统研究. 情报科学, 2003(3)
- 10 张维明等. 信息系统集成技术. 北京: 电子工业出版社, 2002
- 11 Birmingham W, et al. The Distributed Agent Architecture of the University of Michigan Digital Library. IEEE Computer, 1996(5)
- 12 徐彬荣. 多数据源集成新技术. 情报杂志, 2003(1)
- 13 周善儒. 基于 CORBA 中间构件的数字图书馆异构资源集成方法研究. 现代图书情报技术, 2003(2)
- 14, 15 张凤斌. 数字图书馆的集成化管理研究. 情报科学, 2003(2)

马大川 武汉大学信息管理学院副教授。通讯地址: 武汉。邮编 430072。

杨红平 武汉大学信息管理学院硕士研究生。通讯地址同上。

(来稿时间: 2003-10-20)