

马文峰

数字资源整合研究*

摘要 数字资源整合是数字资源优化组合的存在状态,是把各个相对独立的数字资源结合为一个新的有机整体。整合方式有:基于OPAC资源系统的整合;对相关数据库内的数据对象去除重复信息;基于数字图书馆应用系统的整合;在OPAC资源整合系统和数字图书馆资源整合系统之间建立多维度关联。参考文献6。

关键词 数字图书馆 数字资源 信息组织

分类号 G253

ABSTRACT In this paper, the author provides a definition of the integration of digital resources, and discusses several ways of the integration, such as the integration of OPAC-based resources, the one based on library application systems and the one to correlate the former two systems. 6 refs.

KEY WORDS Digital library. Digital resources. Information organization.

CLASS NUMBER G253

1 数字资源整合界说

1.1 含义

“整合”可以理解为由两个或两个以上事物、现

象、过程、属性、关系、信息、能量等在符合一定条件、要求的前提下,融合、聚合或重组成一个较大整体的发展过程及其结果。“整合”的实质就在于涵盖了整合后系统内部的功能和各要素之间的关系。

(2)信息数字化存储,存储量大,体积小,节约空间。

(3)建设见效快,利用方便、自由。凡拥有能够和网络连接的计算机就不必亲自到图书馆。可在任何时间、任何地点方便地检索、查询、浏览网上图书馆的信息资源。

(4)实现网上信息资源共享,信息资源利用率高。对于从事一定学科专业的研究工作者来说,使用本专业虚拟图书馆,无异于拥有自己的专业资料室。而且它的信息资源要比专业资料室广泛得多。一些只在网上发布的信息,惟有通过虚拟图书馆而获得;信息的更新比专业资料室快。虚拟图书馆的信息资源由于能够在更大范围内被用户使用,因此产生的社会效益和经济效益将更高。

(5)信息资源具有系统性和易用性。专业的虚拟图书馆是按照一定的逻辑次序对网站网页的信息进行再组织,此外,它还具备关键词的布尔检索、顺序扫描检索、字段限定等检索功能,提供信息集中网址的直接检索和深度检索。

(6)本课题在研究的过程中,注意总结已有研究成果的经验和教训,建立良好的网络信息资源平台,友好的动态网页界面,即使一般的馆员都能操作使

用,能够及时地进行数据的更新和维护,因而能够最大限度地发挥信息资源的价值。

5 创新虚拟图书馆的意义

本课题在研究领域具有十分重要的创新意义,不仅对网络环境下图书馆丰富的信息资源作用的发挥和图书馆建设有着实践指导意义,对重点学科的建设、对今后图书馆的建设亦提供借鉴意义。该课题的研究,对于高等院校的教学、科研有着十分重要的应用价值。

参考文献

- 1 陈光祚.论“图书情报学虚拟图书馆”的建设.中国图书馆学报,2000(1)
- 2 索传军等.网络环境下图书馆虚拟馆藏的建设.中国图书馆学报,2000(1)
- 3 陈光祚等.试论虚拟图书馆的建设.图书馆杂志,2000(6)
- 4 清华大学虚拟图书馆镜像网站.Calis教育文献保障体系

王桂平 南京师范大学图书馆讲师。通讯地址:南京市。邮编210097。

(来稿时间:2001-12-12)

*本文系中国人民大学2001年科学研究基金项目“研究型大学数字图书馆建设”研究成果之一。

“数字资源”主要指经过一定程度加工整理过的、一个个相对独立的不同类型不同学科的数字资源系统,不包括网上无序的和自身没有控制的数字信息资源。“数字资源整合”是数字资源优化组合的一种存在状态,是依据一定的需要,对各个相对独立的数字资源系统中的数据对象、功能结构及其互动关系进行融合、类聚和重组,重新结合为一个新的有机整体,形成一个效能更好、效率更高的新的数字资源体系。数字资源的整合程度直接关系到它能否被高效吸收与利用。

1.2 理论基础

数字资源整合的理论基础是系统论方法。系统是由相互作用、相互依存的若干组成部分结合而成,具有特定功能的有机整体。系统方法就是按照事物本身的系统性把对象放在系统的形式中加以考察的方法,其显著特点是整体性、关联性和优化性。整体性是系统方法的基本出发点,是指系统诸要素有机集合起来的不同于孤立状态下各要素机械相加的整体性能。关联性是整体性的延续,是指系统的要素之间、要素与系统整体之间的相互联系。优化性是指在一定条件下系统的组织、结构和功能的整体改进,使得系统有机整体具有较大的灵活性和得到发展。

数字资源体系是由各个相互作用、相互联系的数据库整合成的有机整体。以系统论观点,确定资源整合的方法论原则,将系统论方法运用于数字资源整合实践中,从而实现数字资源的整体优化,具有重要现实意义。

1.3 基本原则

整体性原则,是指要保持数字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。

连续性原则,是指数字资源整合的发展性和不间断性。只有通过连续、系统、动态的整合过程,数字资源才能发挥持续效用,才具有生命力。

针对性原则,是指数字资源整合的目的性。整合后的数字资源应满足特定用户需求。

层次性原则,是指数字资源整合的结构性(多维性)。数字资源本身和用户需求的层次性,要求按多种类型、多种层次、多种方式进行多维整合。

科学性原则,是指对数字资源的整合对象、内

容、方式要进行科学论证,切忌随意凑合、拼合。

优化性原则:是指运用一定的技术手段和方法,使数字资源得到合理组合,取得最好的组织结构和组织功能。

1.4 主要特征

(1)整合后的数字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)因是根据学科、专业来整合资源,所以具有较强的专业性。专业性是建立在整体性基础之上,建立在整体性上的专业性是更为优化的专业性。

(3)整合后数据对象间具有统一性和有机关联性,体现了数字资源的系统性。这种关联性反映了学科间的内在联系。

(4)整合后的数字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。

(5)整合后的数字资源在结构功能上的统一性便于数据的更新和扩充。

2 数字资源整合缘起

2.1 数字资源的构成

从CALIS数字资源系统建设情况来看,目前图书馆数字资源主要包括:

(1)自建数据库:包括中外文馆藏书目数据库、中文现刊目次库、学位论文和会议论文数据库、教学参考书数据库以及各类中文特色数据库等。

(2)联机数据库:CALIS引进国外联机数据库20多种,囊括了最主要的品种,类型包括题录、文摘和全文数据库,如OCLC、FirstSearch、EI、EBSCO等。国内联机数据库如中国期刊网、万方数据资源系统、国研报告、中经专网等。

(3)光盘数据库:包括网络光盘数据库和单机光盘数据库。如中国人民大学图书馆购置的人文社科光盘数据库已达100余种。

(4)学术导航库:CALIS 47所院校图书馆共建设了204个重点学科导航库。从整体上说,CALIS数字资源建设涉及面广,几乎涵盖所有的学科和重点专业^[1]。

2.2 数字资源整合的动因

数字资源整合的直接动因是数字化实践发展的需要。从数字资源内容来看,目前存在以下问题:一

是内容交叉重复,影响用户对信息的选择与获取。二是存在冗余信息。多数数据库的资源收录原则是“全”,使得大量使用价值不高的信息进入资源系统,干扰用户对信息的获取,这在人文社科领域尤为明显。三是知识关联程度低。人类知识是一个紧密联系的有机整体,而现有数据资源系统内的数据对象大都是孤立存在的,无法体现学科知识的内在联系。四是目前数字资源系统大多为题录型数据库,二次数字资源和一次数字资源之间缺少链接关系,用户难以获取全文。

从技术角度看,不同的数字资源系统有着不同的编码结构和表达方式,数据格式的不同导致描述和组织标准的差异,导致检索途径和方法的不同;不同的数据库使用不同的检索软件,使得数据库网检索界面也风格各异,迥然不同,具有差异性和复杂性。

图书馆数字资源数量虽呈增长趋势,但因其整体的无序化,内容组织程度不高,数字资源间交叉关联程度较低,用户需要在不同的网络环境之间穿梭漫游,需要在不同的信息空间来回切换,需要掌握不同检索软件的使用方法。从某种意义上讲,数字资源量越大,给用户造成的负担也就越重。如果不对数字资源进行合理有效的整合,必然会使用户陷于不得门径而入的困惑境地,影响着数字资源的有效利用。

3 数字资源整合基本构想

3.1 数字资源整合的立足点

数字资源整合应立足于数字资源系统结构性优化整合。系统的结构性是指系统中各要素之间所形成的某种关系,是系统有机联系的反映。相同的要素具有不同的结构形式时,系统就会产生不同的功能和效果。数字资源体系的结构性是一种组织性的表现,是指数字资源体系中不同数据库之间所形成的某种关联性,是数字资源系统有机联系的反映。

数字资源系统的结构整合包括横向纵向结构整合。横向结构指的是数字资源学科专业上的相互关联性,横向整合即对不同的数据库中相同学科专业的数字对象进行优化整合。纵向结构指数字资源在空间上的相互关联性,纵向整合则是将不同学科专业的数字资源整合为具有多维立体网状结构的有机

知识整体。

结构是功能的基础,结构决定功能。要发挥数字资源系统的整体功能,必须合理地整合结构,构建有效发挥功能的数字资源体系结构。

3.2 数字资源整合方式

它包括多种类型、多种层次、多种方式的整合,目前主要有以下四个层次上整合方式。

(1) 汇合整合方式。主要是基于 OPAC 资源系统的一种整合方式。OPAC 是图书馆重要馆藏数字资源系统,对 OPAC 资源系统的整合是图书馆数字资源最基本的整合方式。

它的实现主要是通过执行 Z39.50 协议,聚合不同平台上的异构 OPAC 数据库,建立书目整合检索系统。用户只要通过一个 OPAC 系统界面即可检索相关图书馆的 OPAC。如上海交通大学图书馆的“Webpac 检索系统”的“多节点数据库检索”,可在统一检索界面上通过著者、题名、主题、关键词、标准书号等途径检索上海交通大学图书馆、华东师范大学图书馆、复旦大学图书馆、上海图书馆的 OPAC 书目信息,不需要在各个图书馆不同的 OPAC 界面间来回切换^[2]。韩国“国立图书馆整合信息系统”可以一次性检索到图书、期刊、CD-ROM、录像、音响等所有资料^[3]。

汇合整合方式实际上是不同 OPAC 系统中数字信息量的综合与合并,是多个馆藏的简单相加,只是通过统一的查询界面,显示各馆 OPAC 系统符合检索条件的命中记录,用户还须逐一点击各馆藏记录,自己进行查重处理,进行二次选择。

(2) 组合整合方式。即对相关数据库内的数据对象去除重复信息的整合方式,提供给用户的不仅是统一的查询界面,而且是不重复和高质量的信息。组合方式是多个数据库系统的有机优化整合,既是不同数字资源系统中信息量的综合,也体现了数字信息质的优化。如“Web of Knowledge 学术资源体系”不仅提供跨数据库交叉检索功能,而且对命中记录做了查重处理,节省了用户对各数据库重复内容进行去重的时间和精力^[4]。

(3) 重组整合方式。重组整合方式主要是基于数字图书馆应用系统的一种资源整合方式。其主要功能是通过将数字资源的分解重组,按数字资源的逻辑关系组织成立体网状、相互联系的知识资源系

统。这是当前数字资源整合的一种新模式。

数字图书馆应用系统是进行数字化建设及整合各类数字资源的基础平台,支持对知识和数字化资源的采集、加工、处理、存储、归档、组织、发布和利用等全过程,一般由数字资源建设体系、数字资源存储与管理体系和用户服务体系三大模块组成。如 IBM 公司和中国文津信息技术研究中心开发的数字图书馆应用系统。

该整合方式的主要优势有:其一,能够整合多种媒体多种渠道的数字信息,不仅包括文字信息,还包括图像、音频、视频等不同载体、不同介质的数字信息;不仅包括数字化的纸质资源,还包括网络各类学术资源。其二,能够实现不同类型、不同级次(一次和二次)资源间的链接,建立起图书、期刊、会议、机构、人物等科学研究所需核心资源和相关资源间的整合与链接关系。其三,能够保持知识体系的整体性和关联性,通过知识因子的有序化和知识关联的网状化,沟通相互隔绝的学科领域,使之成为相互渗透、相互作用的有机体,发挥科学知识的整体功能。其四,能够形成具有新的组织结构和功能的资源系统。这种整合不是简单的“库集合”和“库链接”,而是剔除冗余、重复和劣质信息,对数字资源内容与结构进行新的类聚和重组,形成一个获取便捷、利用率高的新数字资源体系。例如,“中国试验型数字式图书馆”,实现了资源的深层标引和分布式资源库的跨库连接,对分布在不同地点、不同单位制作的不同资源库之间,与书目资源、规范资源、百科全书资源和网上读书之间的连接与查询^[5]。

(4) 一体化综合整合方式。以 MARC 格式描述资源对象的 OPAC 是目前图书馆馆藏主体资源系统,是用户利用数字馆藏的切入点。MARC 注重形式、结构复杂、制作成本高的局限,使其难以承担对大量多媒体数字资源进行编目的重任,也无法完全实现知识资源的立体化和网状化整合。数字图书馆应用系统是以 DC 格式描述资源对象,DC 虽有基于

内容、简单明晰、可修饰可扩展等优点,通过 DC 可以对知识资源进行纵向和横向整合,但不可能解决所有资源的元数据描述问题,还不是十分稳定成熟的元数据标准,也不可能完全代替 MARC。

这种情况下,数字图书馆的资源整合既要考虑图书馆原有的 OPAC 资源基础,又要考虑多种数字资源的建设,一体化综合整合方式便成了图书馆数字资源整合的不二选择。一体化综合整合是指在 OPAC 资源整合系统和数字图书馆资源整合系统之间再建立多维度关联,以形成更为优化的数字图书馆资源整合系统。其中之关键是要实现 OPAC 系统和数字图书馆应用系统之间的技术兼容,实现各种元数据之间以及与其他资源对象之间的互操作。在这方面有许多问题需要探索。上海数字图书馆的以 DC 为核心元素集,多种元数据方法并存,并以基于 XML 结构的 RDF 资源描述体系封装整合多种元数据的设想,为实现综合整合方式提供了一种思路和途径^[6]。

参考文献

- 1 朱强. CALIS 的数字资源建设及发展. 河北科技图苑, 2001(2)
- 2 <http://webpac.lib.sjtu.edu.cn/webpac>
- 3 尹熙昌. 韩国数字图书馆、著作权保护现状. 中国数字图书馆工程知识产权论坛会议论文集(上册), 2002 年 1 月 12 日至 14 日
- 4 范爱红, 姜爱蓉. 基于知识管理的学术资源整合体系——对 Web of Knowledge 的评价. 现代图书情报技术, 2001(6)
- 5 孙承鉴, 刘刚. 中国试验型数字式图书馆的探索与实践. 现代图书情报技术, 2001(6)
- 6 刘伟, 赵亮. 上图数字图书馆元数据方案, <http://www.istis.sh.cn/istis/dlib>

马文峰 中国人民大学图书馆研究馆员。通讯地址:北京市海淀区。邮编 100872。

(来稿时间:2002-02-20)