数字图书馆信息资源整体化组织的实现

吴叶葵

(浙江财经学院 杭州 310012)

摘 要 指出数字图书馆的信息资源有三种来源:自建的本地数据库;自建的网上虚拟资源数据库;引进的数据库。本文提供了一个数字图书馆信息资源整体化组织的实现方案,从整体化组织的规划、整体化组织的实现,到整体化组织的优化,提供了一些建设性的意见。

关键词 数字图书馆 信息资源 信息资源组织

自美国科学家于20世纪90年代初提出数字图书馆这一概念以来,数字图书馆的研究和实践受到世界各国的普遍重视。中国数字图书馆工程于2000年4月正式启动,数字图书馆资源数据库的建设、相关技术的研究、相关标准的制定,是现阶段数字图书馆的主要工作任务。数字图书馆中的信息资源有三种来源:自建的本地数据库;引进的资源数据库;搜集、整理网络信息资源建设的虚拟数据库。在网络环境下,如何利用有限的人力、物力、财力,实现数字图书馆信息资源的整体化组织,更好地为用户服务,是一个值得探讨问题。

1 数字图书馆信息资源整体化组织的规划

信息资源的整体化组织是数字图书馆建设的核心任务,它需要耗费大量的人力、物力、财力。因此,在具体的组织实施前,要进行精心细致的规划,在规划中要明确以下三个方面的问题。

a.根据国家信息资源建设的整体规划要求和用户的信息需求特点,确定要建设什么样的数据库,引进什么样的数据库。中国数字图书馆工程明确规定其建设的原则是:统一规划,制定标准,联合建设,资源共享,防止重复建设。因此,各数字图书馆在进行本地数据建设时,一定要服从国家信息资源整体规划的要求,选择本馆的特色资源数字化后上网,建设特色数据库,搞好与其它馆的分工合作,实现资源的互通有无,共建共享。虚拟数据库的建设,要根据本地区经济建设、学科建设的要求,组建特定主题的专题指引库。此外,还要根据本地用户的信息需求特点和本馆的经济实力,有选择地引进数据库。

b.确定要采用什么样的技术路线。数字图书馆是建立在先进计算机技术、网络技术、信息处理技术基础上的高新技术项目。如今,这些技术的发展可谓日新月异。因此,中国数字图书馆工程规定要采用与国际同类主流技术有接轨前景的技术方案。具体来说,要采用通用的置标语言(SGML /XML)、统一资源名称(URN)等;严格遵循电子信息处理与电子信息交换的相关国际标准及工业标准;采用适用于网络环境的分布式面向对象的软件技术;立足于自行开发和引进成熟技术相结合。在相关标准的制定和应用中,应既考虑与国际标准接轨,又兼顾自己的资源特色,实现标准的实用性。

c.人力、物力、财力的规划。数字图书馆是一项耗资巨大的系统工程。根据国外的经验,数字图书馆的建设,通常由国家政府提供启动资金,由民间组织提供赞助,或自筹资金,滚动发展,走公益性和商业化相结合的道路。这种方法值得我国借鉴。数字图书馆的建设需要多方面的人才,包括行政管理人员、财务管理人员、计算机网络管理人员、图书情报专家及一般的操作人员。各种人员需要多少,任务怎样分配,需要根据数字图书馆建设的任务、目标、难度和进度要求来定。

2 数字图书馆信息资源整体化组织的实施

数字图书馆信息资源整体化组织的实现是一个长期的、复杂的、任务艰巨的大工程,需要各方面的人才同心协力,需要相关的建设单位鼎力相助。数字图书馆信息资源整体化组织的实现包括三方面的工作:本地数据库的建设、虚拟数据库的建设、数据库的引进。

2.1 *本地数据库建设* 在数字图书馆信息资源整体化组织中,本地数据库的建设是最重要的工作。只有拥有有特色的本地数据库,才能成为有特色的数据图书馆。本地数据库的建设,是数字图书馆根据信息资源共建共享的目标,根据本馆的资源特色和自己的经济实力,选择有价值的资源数字化后上网,建设有特色信息资源库。数字图书馆本地数据库的建设分以下几个步骤:

a.为资源库的建设编写脚本。在这一步里,要根据本地数据库建设的目标进行信息资源的搜集、整理;为提高信息资源的表现形式,有时还需对搜集到的音、视频素材进行编辑;然后,根据搜集到的素材,确定元数据格式,提出数据结构要求,供软件人员设计文献类型定义。

b.资源内容的再制作。对非数字化资源进行数字化转换,然后选择合适的格式压缩,并将其存储在海量的存储器里。

c.对资源内容的标引。由标引人员对经过再制作的数字文件进行标引,包括分析内容、给出主题分类,并使用基于 SGML / XML 开发的资源加工系统软件对资源内容置标。

d.质量检查和归档。对加工后的文件进行质量的检查,包括检查声音质量、图像大小、图像质量以及标引的正确性等方面。对于经检查合格的文件,将其归档,存入资源库。

e.元数据的抽取。由人工或计算机软件自动抽取元数据,建立索引数据库。

2.2 *数据库引进* 数据库的建设要耗费大量的资金,因此数据库的引进大都价格昂贵。中国高等教育文献保障系统 CALIS,

规定由国家中心统一规划，引进国外数据库，供各地区中心共同使用，所以在CALIS各中心的网页上，可以经常看到更新的引进数据库。数字图书馆数据库的引进工作，要服从全国范围内数字图书馆资源建设统一规划的要求；要考虑用户的信息需求，不能盲目引进；要考虑本馆的经济实力，力争以最少的资金，实现最大的经济价值和社会价值。

2.3 *虚拟数据库的组织* 信息时代，因特网上信息资源呈爆炸的趋势在增长，用户的需求日新月异，不断更新。如果要将用户所需的信息一一纳入到数字图书馆信息资源库中，是不可能完成的工作。因此，针对本地经济建设或学科建设中特别重要的主题，组织网上虚拟数据库，是现实可行的途径。所谓虚拟数据库，是指按学科、专业、主题建立起来的网络信息资源指引库，在数字图书馆的本地一般只存储这些信息的索引数据和URL地址，而原始信息则广泛地分布在网络各地。当然，对于用户经常使用、有较高价值、又有办法解决知识产权问题的信息，也可以下载到本地存储，以降低通信费用和提高用户的访问效率。但虚拟数据库建设的主要目的不是存储原始信息，而是把Internet上与某一或某些主题相关的节点进行集中，按照方便用户检索的原则，用用户熟悉的语言组织起来，向用户提供这些资源的分布情况，指引用户查找。虚拟数据的组织包括以下步骤：a.虚拟数据库主题树体系结构的设计；b.人工或利用计算机软件采集网上信息；c.对信息资源进行评价和筛选；d.对筛选后的信息进行加工组织，包括资源描述、资源标签、资源排序，最后为网络资源建立著录款目文档和索引文档。

3 数字图书馆信息资源整体化组织的优化

数字图书馆信息资源的组织不是信息资源的简单堆砌，对于入库的所有信息资源，要从用户需求的角度出发，对其进行整体优化，以最方便、最快捷的方式向用户提供所需信息。数字图书馆信息资源整体化组织的优化可从三个方面入手。

a.建立标准友好的检索界面，提供多种检索途径。数字图书馆检索界面必须遵循一定的标准。对于声音、图像、视频、文本等不同类型的数据，数字图书馆应提供不同的检索模式，但对于相同类型的数据资料，数字图书馆在检索命令、检索符、检索表达式的构造方面，应与其它的数字图书馆遵循统一规范。数字图书馆的检索界面必须简洁、庄重、通俗易懂、画面漂亮大方，并力求友好，尊重用户的思维方法和思维习惯，减少系统对用户的限制，能让用户尽可能真实地表达自己的信息需求。用户在检索中碰到问题时，系统须及时友好地给予帮助。此外，数字图书馆应利用先进的信息处理技术对信息资源进行多维的揭示，向用户提供多种检索途径。如对于文本信息资源，应能提供书名(篇名)、作者、主题词、分类号、期刊名、出版机构、引文等检索功能外，尽可能地提供全文检索功能；对于音、视频信息，除提供对其标引著录项的检索外，还应能提供基于内容的检索。

b.整理入库的信息资源，提供针对不同专题的推送服务。传统的信息检索服务是一种"拉"式的服务。检索的过程是：用户提供检索词，检索系统利用检索词检索数据库，得到相关信息的URL地址，检索系统将URL地址返回给用户，用户再利用URL地址查询相关网站，然后将相关信息"拉"回本地客户机。信息服务的发展方向是基于人工智能的主动推送服务，也就是由用户向系统提供(或由人工智能根据用户利用信息的情况自动归纳)所需信息的相关主题，以后系统自动地将用户所需信息推送到用户的客户机中。目前，人工智能正在发展之中，基于人工智能的推送服务，也为一些信息资源组织机构所尝试。数字图书馆，作为一种先进的信息资源组织机构，应有选择地提供这种推送服务。数字图书馆应选择一些人们关心的热门话题，或与学科建设、经济建设相关重点问题，整理入库的信息资源，提供主动的推送服务。

c.对于不同数据库中的信息资源，提供统一的检索入口。进入数字图书馆资源库的信息有不同的来源，有传统的机读目录，有经扫描录入、数字化后的文本信息，有经数字化转换的音、视频资料，有经整理入库的网上虚拟信息，有引进的数据库资源，这些以不同的元数据格式存储在不同软、硬件环境的数据库中。从用户检索的角度出发，用户不希望一个数据库、一个数据库的检索，而希望系统能提供统一的检索入口，一次检索得到所需的全部信息，因此数字图书馆有必要对入库的各种信息资源进行整合。

对不同元数据格式的信息进行整合，需要解决不同元数据之间的互操作问题。目前，解决元数据之间的互操作有三种方案：a.元数据映射，即实现元数据格式的转换。如今，已有大量的转换程序存在，提供若干种元数据格式之间的转换。b.标准描述方法。即提供一种标准的资源描述框架，用这个框架来描述各种元数据格式，那么，一个系统只要能解析这个框架，就能解读所有的元数据格式。W3C制定的RDF(Resource Description Framework)就是这样一个资源描述框架，它通过资源、属性、声明三种对象构成的简单模型，为各种元数据提供了一个容器，从而实现各种元数据的解读。c.数字对象的方法。即建立包含元数据的数字对象，在该数字对象中定义元数据格式转换机制，实现该种元数据向其它元数据格式的转换。

Stanford数字图书馆计划的元数据资料构架，采用元数据映射的方法来整合不同元素据格式的信息资源。在此构架中，用属性模型代理(Attribute Model Proxy)来表示真实世界里的各种元数据。通过属性模型转译服务(Attribute Model Translation Service)实现各种元数据之间的转换。转换后的元数据存放在统一的元数据库(Metadata Repository)中，在统一元数据库的基础上，提供统一窗口为用户提供检索服务。

目前，在我国数字图书馆信息资源的建设中，信息资源的整合问题还没得到足够的重视，用户进入数字图书馆，仍需针对一个一个的数据库进行检索。这方面的工作以后还得加强。

参 考 文 献

1 中国数字图书馆工程：http://www.d-library.com.cn/go/index.htm

2 高 文，刘 峰，黄铁军等．数字图书馆原理与技术实现．北京：清华大学出版社，2000

3 http://www.calis.edu.cn/

4 陈梅华．探索网络信息资源建设的关键技术——建立指引库和自动跟踪．情报学报，1997；(2)

5 M.Baldonado，C-Ck.Chang，L.Gravano and A.Paepcke.The Stanford Digital Library Metadata Architecture.International Journal on Digital Libraries，1999；(1)

6 柯皓仁，黄凤贤，杨维邦．元数据与数字图书馆系统互通性之探讨．大学图书馆

(责编：王京梅)