

21 世纪的数字富翁

数字化

SHUZHUA

TUSHU YU

SHUZITUSHUGUAN YINGYONG YANJIU

图书与数字图书馆应用研究

王建文 编著

北京工业大学出版社

21 世纪的数字富翁

数字化

SHUZHUA

TUSHU YU

SHUZITUSHUGUAN YINGYONG YANJIU

图书与数字图书馆应用研究

王建文 编著

200813

北京工业大学出版社

图书在版编目 (CIP) 数据

数字化图书与数字图书馆应用研究 / 王建文编著.
北京: 北京工业大学出版社, 2005.7
ISBN 7-5639-1553-2

I. 数... II. 王... III. 数字图书馆 - 应用 - 研究
IV. G250.76

中国版本图书馆 CIP 数据核字 (2005) 第 067169 号

数字化图书与数字图书馆应用研究

编 著: 王建文

出版发行: 北京工业大学出版社

地 址: 北京市朝阳区平乐园 100 号

电 话: (010) 67391474 (010) 67391106

网 址: <http://www.cctp.com.cn>

E-mail: edit@cctpbook.com

经 销: 全国新华书店

印 刷: 湖南版艺印刷有限公司

开 本: 787×1092 1/16

字 数: 273.6 千字

印 张: 13

版 次: 2005 年 7 月第 1 版 2005 年 7 月第 1 次印刷

ISBN 7-5639-1553-2/G·789

定 价: 25.00 元

前 言

在任何时候、任何地方能否快速找到您所需要的知识信息？通过阅读本书，回答是肯定的！数字图书馆作为人们方便快捷地获取、利用知识的有效工具，已引起全人类广泛的关注。利用它，人们可以通过互联网实现对人类知识的普遍存取；利用它，人们可以不受时间和空间限制，通过任何与互联网相连的设备，轻点鼠标就可获得自己所需要的一切信息。尤其是作为学校三大板块之一的图书馆，都纷纷把图书馆数字化建设视作追捧的目标，教育教学资源和学习资源的无限制地广泛地共享与交流，一直是广大师生梦寐以求的愿望，只有利用数字图书馆的先进信息技术平台，这一美好的愿望才得以真正实现。那么，数字图书馆到底是怎样的一种图书馆？它的组成结构及实现原理如何？当前实现其功能的先进方法与技术及其研究方向如何？它在现代信息社会中为何富有如此诱人的吸引力？怎样借助这一神奇的信息技术工具为我所用？带着这些问题，请翻开本书，我们来一页一页地揭开数字图书馆的神秘面纱。

数字图书馆对于传统图书馆来说，是一个全新的理念，也是一种全新的实践，国内外可供参考的文献并不很多，特别是结合中国数字图书馆建设的实际经验加以深入阐述的文献很少。本书的编写以国内外数字图书馆研究和建设的最新成果为基础，以相关信息技术为主线，力求突出理论的严谨性、技术的先进性、方法的系统性。全书内容体现了数字图书馆理论、技术、应用三大组成部分，力求全面揭示数字图书馆的整体构架及核心要素。本书尽量做到立足当前，面向

未来,资料翔实,深入浅出,可读性较强。数字图书馆具有无限发展的空间,其理论与实践都在不断更新、完善,希望本书在促进数字图书馆基础知识的普及和数字图书馆理论与实践的探讨方面能够充当引玉之砖。

本书的编写得到了湖南师范大学图书馆领导和有关图书情报专家的关心指导和大力支持,得到了中南大学数学与计算机软件学院教授蒋德顺及中南林学院计算机系刘军万教授的悉心指教,在此表示衷心的感谢。

在本书的编写过程中,编者参考了国内外大量文献,在此谨向参考文献的作者表示深深的感谢!

由于编者水平有限,时间紧迫,本书虽经多次修改,其中错漏之处仍在所难免,敬请广大读者及专家不吝指正。

编 者

2005年5月26日



目 录

第一章 概 述	1
一、数字图书馆的简要历史	1
二、电子图书馆、虚拟图书馆与数字图书馆	2
三、数字图书馆与传统图书馆的关系	3
四、数字图书馆产生的背景	4
五、数字图书馆的概念	6
六、数字图书馆的特征	9
七、国内外数字图书馆的发展概况	12
第二章 数字图书馆的功能与价值	16
一、信息数字化和信息搜集	16
二、信息的获取与创建	16
三、信息的存储和管理	16
四、信息服务	17
五、信息传播	18
六、信息安全和权限管理	18
七、信息的访问与查询	18
八、数字图书馆的发展功能	18
九、因特网资源的连接功能	18
十、电子出版物资源的利用	19
十一、网络服务功能	19
第三章 图书馆数字化是实现数字图书馆的基础	20
一、数字化的概念	20
二、数字化是实现数字图书馆的前提	21
三、数字化资源的建设	22

四、数字化技术	26
五、数字化的过程与方法	32
第四章 实现数字图书馆的关键技术	35
一、数字图书馆的核心技术简介	35
二、网络信息资源的组织管理技术	60
三、便捷的导引系统构建及实现方法	71
四、数字化信息资源标准化技术	78
五、数字信息传递技术	80
六、信息检索技术	81
七、超文本技术	103
八、数据仓库与信息挖掘技术	110
第五章 数字图书馆的体系结构	117
一、数字图书馆的基本构件及系统构成	117
二、数字图书馆的信息结构	118
三、数字图书馆的系统结构	123
四、IBM 数字图书馆系统简介	128
第六章 数字图书馆的未来前景展望	132
一、美国数字图书馆的研究与发展概览	132
二、我国数字图书馆的发展思路	134
三、我国数字图书馆建设的必要性与紧迫性	136
四、国际性超大规模数字图书馆的原型 及其挑战性问题研究	138
第七章 怎样利用数字图书馆	144
一、数字图书馆信息服务及服务的组织管理	144
二、网络信息资源导航与利用	145
三、网络环境下数字图书馆的信息服务	150
四、数字图书馆数字化参考咨询服务	152

五、利用网络开展虚拟咨询	156
六、网络环境下的数字图书馆开展 个性化的信息服务	159
七、国外数字图书馆 Mylibrary 个性化服务 系统简介	162
八、利用电子邮件开展信息服务	164
九、数字图书馆门户网站的信息功能	168
十、数字图书馆的读者教育	169
十一、当前数字图书馆用户的信息素质教育	171
第八章 数字图书馆在教育领域的功用	174
一、构筑数字化教育信息资源中心	174
二、构设先进的网络教学平台	175
三、构造方便、个性化的学习环境	177
四、实现教育教学资源的交流与共享	178
五、数字图书馆网络教育信息资源的利用	178
参考文献	196

第一章 概 述

以计算机技术、网络通信技术为代表的因特网迅速崛起,推动人类社会在经历了农业社会和工业社会后,开始进入信息社会。与此相适应,图书馆在走过传统图书馆阶段、自动化图书馆阶段后,也开始步入数字图书馆阶段。这将是图书馆发展历史上的一次变革。数字图书馆自 20 世纪 90 年代初在美国问世以来,经过十余年发展已经形成难以计数的项目和成果。数字图书馆的概念和应用逐渐地深入人心,它已经成为人们日常工作、学习不可或缺的有效工具。有关数字图书馆的文章充斥了各种专业期刊,在教育科研领域,尤其在图书信息情报行业这是一个耳熟能详的名字。数字图书馆正在全世界迅速发展,它对传统的图书馆的未来发展带来了无可限量的发展机遇,同时也提出了一系列严峻的挑战。因此,数字图书馆是图书情报工作人员、教育与计算机科研人员以及想借助数字图书馆开展教学教研活动的广大教师值得关注和研究的新领域。

一、数字图书馆的简要历史

真正最早应用数字图书馆概念是在 20 世纪 60 年代。当时,人们开始以全文存储的格式,对法律和科学文档进行索引。使用 OBAR (Ohio Bar Automated Research) 来提供对法律资料的联机访问功能。美国空军研制的 LITE (Legal Information Through Electronics) 系统可以对所有法律文件进行索引。但是,这些早期的文本存储和检索技术都是针对特定文档及其格式的,以专用编程方法实现,不便普遍推广。20 世纪 70 年代出现了通用的基于计算机的信息存储、索引和检索程序。最著名的一个例子是 IBM 存储和信息检索系统 STAIRS (Storage And Information Retrieval System), 原来是为大型机设计的,后来它们被数以百计的图书馆采用,用于图书馆流通管理。

随着计算机技术的广泛应用,图书馆管理本身也在计算机化,这就是图书馆信息管理系统。典型的图书馆信息管理系统具有采编管理、编目管理、流通典藏管理、期刊管理和公共检索的功能,能够对图书预定、验收,图书编目,图书借还、预约、续借,典藏分布,期刊订阅进行计算机管理,也逐步提供远程的书目检索。

图书馆是最早采用信息检索系统的场所之一,并在信息检索技术的支持下,实现了联机书目系统。早期的检索系统种类多,对存档资料进行检索的典型形式有两种:搜索远程的电子数据库,搜索图书馆内部的资料目录。

当时这两种类型的系统是独立发展的,我们把第一种系统简称为远程数据库检索,后一种系统称为联机公共访问目录 OPAC(Online Public Access Catalogs)。它们之间的主要区别在三个方面:系统开发的起源、文件和数据库的内容、用户。

远程数据库检索系统是以科研项目支持的形式,由政府的实验室承担发展起来的,它基于书目数据库,主要包含文本信息,通过受过训练的中间人来查询。而 OPAC 是由图书馆开发的系统,它一般采用标准化的记录格式(例如机读目录 MARC 格式),包含主题信息,例如书名、主题、分类号等。不像商业的远程数据库检索系统,它是为最终用户而设计的。由于这些设计和考虑的不同,使得这两种系统在使用方式和特性等方面都有较大的不同。

然而,这些联机书目系统仍然有它们的限制,因为它们利用的仅仅是文本字符数据,没有文档的图像。虽然文献资料主要包含的是文字,但是作为图书馆资料,具有版面的“图像”,包括文字、图形和照片,从这个意义上来说,具有表现样式的多媒体文档存储和检索能力是数字图书馆的一个要素。因此需要利用多媒体信息技术,把文档资料以多媒体格式而不仅仅是文字格式来存储和被检索。

随着技术的进步,数字图书馆的概念也在变化。1991 年美国计算机学会 ACM 和国家科学基金会 NSF 联合支持的计算机科学文献的数字图书馆原型系统 Envision,它强调可用性(用户为中心的设计)、高度结构化(SGML 和对象数据库)和高度集成性(建立对象之间的超文本链)。

同年(1991 年),美国化学学会、化学文摘、康奈尔大学等其他组织共同支持了电子图书馆原型 CORE(Chemistry Online Retrieval Experiment)的建立,它保藏的是化学方面的资料,每篇文摘包括扫描的图像和 SGML 文档,系统提供全文形式的布尔搜索。

1994 年,美国国会图书馆宣布国家数字图书馆 NDL(National Digital Library)计划,利用现代信息技术,管理馆藏中的印刷和非印刷的资料。NDL 与其他组织合作,希望能够建立一个广泛的与其他图书馆资源相通的大系统,并作为国家信息基础设施整体中的一部分。

同时在 1994 年,美国 NSF、NASA 和 DARPA 联合启动了数字图书馆预研计划 DLI,支持六所大学的数字图书馆研究。在 1994—1999 年期间,投资超过 6 800 万美元。目前正在进行的是第二阶段的计划 DLI-2。

数字图书馆的历史不断发展和变化,从数据库的利用和图书馆的电子化,发展到现在的无所不包的数字图书馆 DLI-2 目标,已经远远超越了传统图书馆的概念。

二、电子图书馆、虚拟图书馆与数字图书馆

数字图书馆、电子图书馆、虚拟图书馆(无围墙图书馆)等概念之间有细微的差别,但由

于它们内涵相似,因而目前普遍将其视为近义词,在很多情况下可互相替代使用。

电子图书馆侧重对藏书对象、特色的概括,即强调信息资源的表现和存储形式,尤其是信息的载体形式。电子图书馆的收藏品主要是电子读物,其信息资源的主要形式是磁带、磁盘、CD、ROM 或缩微胶片,以模拟信号或光信号录制的音像制品等。电子化的图书馆并不排斥在传统的管理模式下工作,读者通过计算机进行阅读,不一定提供网络信息或服务。

无围墙图书馆,即“虚拟图书馆”是一种网络环境下的信息资源管理与服务机构。人们在网络上漫游、查询和检索信息,就好像是在一个图书馆里。虚拟图书馆是将网上信息进行组织,供读者查询与阅读,不一定需要对应的图书馆实体。它可以看成是数字图书馆的初级形态。虚拟图书馆侧重图书馆无实体的特征,它形象地说明了“数字图书馆”非图书馆实体的特征。

在数字图书馆概念出现之前,人们从电子图书馆、无围墙图书馆和数字化图书馆等不同侧面讨论了图书馆的未来发展。电子图书馆主要是指以电子形式(媒介)存储、传递并提供服务的图书馆,如各个图书馆建立的电子阅览室、OPAC 检索等,它重点强调个别图书馆借助于计算机技术在一定范围内提供读者服务。无围墙图书馆也称虚拟图书馆,是指读者通过网络通信系统检索和利用多个图书馆的数字化馆藏资料,强调用户获取信息的广泛性和网络传输的重要性。数字化图书馆则主要侧重于馆藏资料的数字化,将数字化的书刊文献上网提供给读者使用。可以看出,电子图书馆主要侧重于在个别图书馆的电子化检索;无围墙图书馆主要强调通过网络检索多个图书馆的馆藏资料;而数字化图书馆的重点则在馆藏资料的数字化。电子图书馆、无围墙图书馆和数字化图书馆在建设难度、使用的方便程度上都远远不如数字图书馆,但它们是数字图书馆的基础,是建设数字图书馆必不可少的前期准备。

三、数字图书馆与传统图书馆的关系

传统图书馆一般是指具有固定的馆舍、馆藏以印刷型书刊资料为主,服务以图书馆为中心,读者要借阅书刊、查阅资料,需要到图书馆来。数字图书馆与传统图书馆的关系是既有区别又有联系,主要体现在以下几个方面:

(一)数字图书馆是传统图书馆的发展

传统图书馆的藏书以纸质印刷书刊资料为主,通过卡片目录反映馆藏信息,通过读者到馆借阅和送书上门传递信息,整个工作以手工操作为主。数字图书馆则在许多方面有别于传统图书馆,即信息资源由过去的“以藏为主”变成了“藏用结合,以用为主”;信息服务由被动服务、单一服务变成了主动服务、综合服务;书目管理由卡片式变成计算机管理;读者借阅

由封闭型变成开放型;信息资料的贮存由单一的印刷型变成纸质图书、缩微、电子出版物和网络贮存相结合的多元化方式。

(二) 传统图书馆是数字图书馆的基础

数字图书馆作为一种信息环境,离不开各个具体图书馆和信息机构的支持。它的信息资源除了一部分由信息生产者以电子形式生产并提供上网外,大部分信息资源(特别是非电子型的信息资源)还必须靠具体的图书馆或信息服务机构去搜集、加工,然后转变成数字形式才能提供网上利用。

(三) 数字图书馆和传统图书馆将长期共存,优势互补

现有图书馆已积累了大量信息资源,要将它们转换为数字形式提供网上利用,不仅工作量大,而且短时间内无法实现。虽然数字图书馆和传统图书馆在存在形态与藏书规模方面存在明显差异,服务手段与功能也迥然不同,但形式与服务功能的变化并不改变传统图书馆的性质,也无碍传统图书馆的生存。只要社会存在信息,人们需要信息,传统图书馆就不会消亡。

四、数字图书馆产生的背景

图书馆的发展和演变,离不开内部环境的驱动力和社会环境的推动力。内部环境主要是指社会对图书馆的需求产生的驱动力,促使其改变传统图书馆的服务环境,实现图书馆的自动化;社会环境的变化主要包括社会经济结构、信息技术结构、文化结构的变动,也推动了图书馆的发展和演化。

(一) 图书馆内部环境的变革

1969年,美国国会图书馆正式发行MARC II型的机读目录以来,使图书馆走向了标准化、规范化生产机读目录的新时期,产生了图书馆自动化集成系统,代替了传统图书馆的手工作业和业务流程,改善了图书馆的读者服务的手段,提高了服务效果和工作效率。20世纪70年代后,图书馆界、情报界又出现了一大批联机编目中心和大型文摘与检索系统,如OCLC、WLN、RLIN和DIALOG、ORBIT、MEDLINE、BRS等,这些系统较深层次揭示了馆藏图书、期刊的内容,生成了大规模的书目数据库。图书馆已进入书目数字化的阶段,也标志着图书馆有能力推动图书馆自动化系统的发展。进入80年代以后,特别是局域网络技术迅猛

发展与成熟,图书馆建立了以局域网为基础的网络系统,通过局域网络与广域网络连接,实现了远程查询图书馆的机读目录和大型文献信息数据库,这一时期,为文献数字化的网络查询、检索、传输打下了良好的基础。20世纪90年代以来,世界先进国家的图书馆已经完成了图书馆自动化——书目时代的任务,正加紧迈向数字化图书馆的建设,进行大容量、多媒体信息处理平台建设和文献信息资源的数字化转换工作。如美国国会图书馆,1995年进行国家数字图书馆项目建设,仅至2002年就完成了1000万件文献图片的数字化转换处理;我国国家图书馆到2003年底就已完成9000万页文献资料的数字化制作和上网。这一时期,数字图书馆的试验、研制取得了许多重大成果,如快速扫描与压缩技术,图像、声音的采集存贮与检索技术,文献的标识与识别技术,版权保护技术等等。图书馆采用信息技术,实现了卡片目录、书目索引、文摘的数字化,正进入对图书文献内容的全文数字化转换阶段。由于信息技术的应用逐渐改变了图书馆的内部环境,逐渐适应了社会对图书馆的需求。

(二) 社会环境的变革

1. 社会经济结构的变动

在现代经济结构中,信息被赋予为资本,信息资源成为战略资源,产生了新的经济理论——知识经济。另一方面,由于信息技术的迅速发展,产生了新兴的信息产业,而且迅速成为高新技术的主导产业,其中信息有序化处理、信息服务业已成为信息产业中一个重要的分支产业,知识经济为图书馆的发展提供了一个很好的发展机遇,加快了图书馆信息服务自动化、网络化的进程。

2. 技术结构的变革

在当代信息技术结构中,计算机、通信、网络、数字技术已成为当代社会变革的最大的技术支柱,这些技术的成熟与发展,也为图书馆的自动化、网络化、数字化的研制、试验、运行提供了技术基础。在这些技术的支持下,图书馆建成了一系列的高质量的信息服务系统、大型的数据库系统、网络信息系统。随着电子出版物——光盘及光盘网络的应用,因特网的普及与推广,为数字图书馆的产生提供了实实在在的网络环境,为虚拟图书馆提供了一个巨大的信息资源宝库,为网上图书馆提供了WWW环境,为图书馆的信息资源开发与利用提供了样板。信息技术将促进图书馆在更广泛的时空中,向网络化、数字化方向发展。

3. 文化结构变动

在当代文化结构中,逐渐形成了文化、教育、科技三位一体的文化,称为大文化。信息技术改变了文化艺术的表现形式,出现了以数字冠名的设备、影视、多维活动图像、音乐等。如数字照相机、数字摄像机、数字电视、数字影视片、数字音像、数字报刊等,这些都属于电子信息文化的组成部分。这种信息服务必然是数字图书馆才能做到的,而且也是现代图书馆信息服务业的信息增值和经济增长点之一。

4. 信息资源已开始全面电子化、数字化

信息内容已开始全面数字化,因特网资源随着越来越多的数据库和信息系统不断加入而不断增加。就我国公共图书馆而言,目前制作数字图书资料的有中国国家图书馆、上海图书馆、辽宁图书馆、广东中山图书馆、深圳图书馆等。清华大学出版的中国学术期刊(光盘版)已收录国内重要学术期刊2600多种,论文约40余万篇(每年),我国的《人民日报》等20多种重要的报纸都可在网上查阅。所以说,因特网是一座巨大的数字图书馆,也是一座全球的虚拟图书馆。

五、数字图书馆的概念

(一)数字图书馆概念的理解

1994年后数字图书馆的基本概念已确立,“数字图书馆是一种有纸化的(Paper-based)图书馆外观和感觉的图书馆,但这类图书馆资料都已被数字化并存贮起来,而且能在网络环境中被本地和远程用户存取,还能通过复杂和一体化的自动控制系统为用户提供先进的、自动化的电子服务。”数字图书馆仍然是传统图书馆的继续,其不同之处是文献被数字化,并可通过网络存取,还能通过一体化控制达到管理和规范。

有学者认为:“数字图书馆是一个数字化信息系统。它将把分散于不同载体、不同地理位置的信息资源以数字化的形式存贮,以网络化的方式互相连接,提供及时利用,实现资源共享。”这种观点界定了数字图书馆是一个数字化信息系统。

“数字图书馆是计算机数据仓库(Repository)或数字化对象的集合”。这个观点从宏观上界定了数字图书馆是一个数据库系统。

“所谓数字图书馆是这样一种图书馆,其全部或大部分馆藏都是以计算机可处理的形式存在的,以此来补充并逐渐代替构成当前图书馆馆藏主体的传统印本文献或缩微资料”。这种观点只局限于把传统图书馆的馆藏进行数字化的转换。

“数字图书馆是传统图书馆的扩展,既具有传统图书馆收藏、保护、提供服务功能,又提供集成化的、深层次的、数字媒体的远程数字图书馆服务”。这个观点认为数字图书馆是传统图书馆功能的补充,只是馆藏范围、服务方式上的深化和网络化。

归纳上述数字图书馆的定义,虽然用词不同,但都有共同点:馆藏文献的数字化存贮,可通过网络查询和传输;不同点:数字图书馆是传统图书馆功能的扩展,数字图书馆是数字对象的集合,而且具备了自动调度的功能。为了比较清晰地描述数字图书馆的概念与定义,我们不妨从技术角度和图书馆角度去解释数字图书馆的概念与定义。

从技术角度来定义:数字图书馆是能将多媒体的文献信息内容,以数字化存贮与管理,

通过调度系统标识数字对象,通过网络浏览、查询、检索和传输各种信息,采用权限管理技术保护知识产权,通过 Web 发布数字化信息的网上图书馆。数字图书馆是以计算机技术、网络通信技术、数字技术和文献处理技术为基础的大型的数字化的文献资源信息系统。

数字图书馆概念的技术基础:它包含图书馆自动化集成系统的接口与连接技术、数字获取技术、数字描述技术、数字对象标识与对象库技术、关联索引与检索技术、用户界面设计技术(交互式智能化)、版权管理技术和文献处理技术。

从图书馆角度来定义:数字图书馆是传统图书馆功能的扩展,是既能承接传统图书馆的自动化系统,又能使图书馆用户联机存取本地和远程数字化图书馆资源的现代化图书馆。所以,数字图书馆包含:

(1) 数字图书馆仍然是一个能感觉到的图书馆,必然具有图书馆的文献收集、加工、整理、保存、服务的基本功能;

(2) 数字图书馆以计算机可处理的数字形式存贮文献信息,而不同于传统图书馆的多载体的文献信息收藏;

(3) 数字图书馆的数字化信息的收藏范围从广泛性和深层性上远远超出传统图书馆,所以它是以文献资源内容为基础的系统;

(4) 数字图书馆提供更广泛、迅速、便利、多种形式的服务,依托因特网,利用先进的数字处理技术和网络工作站为全球用户提供远程服务;

(5) 数字图书馆是传统图书馆的补充,尽管有人预言数字图书馆将来可以代替传统图书馆,但实际上数字图书馆只在馆藏范围、服务方式上对传统图书馆加以补充与深化,完全代替是不可能的。

因此,数字图书馆的着眼点应在于图书馆馆藏数字化的建设,数字化信息服务的开展,逐渐实现网上资源共享才是最根本的。

1993 年以来,随着数字技术实用化,包括数字视频和音频技术的发展,特别是 IBM 公司推出“IBM 数字图书馆”计划后,许多学者、图书馆员普遍使用数字图书馆这一名词。数字图书馆就是图书馆利用数字图书馆的技术,将其馆藏文献数字化并提供有效的服务和管理,同时利用该技术,组织网上所有可用的数字化信息的链接和导航。这样,用户可以获得图书馆经过数字化的所有文献信息,包括联机采购、编目、公共查询;组织用户通过网络检索本馆的信息资源,访问外界数字图书馆和文献信息数据库系统,如电子期刊、电子书、声像资料、动画片、影视片 and 多媒体资料等;链接图书馆自动化系统,继续管理图书馆的采、编、流、期刊等的服务工作;利用数字图书馆技术建立数字图书馆的网络环境和数字图书馆平台,链接到单位、团体、学校、办公室、工商机构、家庭,让人们能方便地共享图书馆及其他资源;利用数字图书馆技术,办好网上图书馆、网上专题服务、网上公告;利用数字图书馆技术,管理和保护知识产权。

因此从资源的种类看,可以把数字图书馆看成是分布式多媒体信息系统,其关注于数字

图书馆保藏的各种类型的多媒体信息资源:文本、图像、声音、静态和动态图像,并且系统的结构是分布的,提供分布的信息存储和分布的信息存取能力。

从分布的角度看,可以把数字图书馆看成是联邦(分布式异构型)结构的信息库,它为用户提供对大型的不增加的多媒体信息的存取。

从信息检索的角度看,数字图书馆可以看成是信息检索系统的扩展。其扩展主要是在联邦关联和媒体种类两个方面,即对分布的大量的信息库进行检索,检索的对象不仅仅是文本,而且包含对多媒体数据搜索、分类和索引。因此,信息检索系统所面临的是多语种处理、多媒体内容检索、信息可视化、搜索大型分布文档集、可用性、标准、体系结构等方面的问题。

(二)数字图书馆的准则——什么是数字图书馆,什么不是数字图书馆

数字图书馆应具备若干准则,作为衡量数字图书馆的基本原则。

1. 资源的数字化

数字化资源是数字图书馆的基本特征。这些资源包括文本、图形、图像、声音、视频等多种类型的数据,同时,随着数字图书馆的发展,其资源类型会不断丰富和完善,其中的文本数据不仅包括简单的纯文本,还包括如公式、表格等具有一定特殊格式的数据。数字化资源以其特有的格式存储于磁介质或光介质中,如文本以编码方式存储,图形以矢量方式存储,图像、声音、视频以各自的数据格式存储。这些存储形式应具有可检索、可压缩和可索引的特性,能进行识别、深度加工和处理。

以下列资源类型为主的信息系统不是数字图书馆,至少不是真正的、最终意义上的数字图书馆:

- ①单纯的文本数据;
- ②以常规方式转换生成的,未进一步处理的扫描图像数据;
- ③以数字化信息索引,以模拟信号形式存储的声音、视频数据;
- ④由简单的模—数转换生成的,但不能进行计算机自动索引、局部识别、检索匹配的数字化形态声音、视频数据;
- ⑤以数字化信息索引,以印刷形式保存的文献数据。

2. 资源的规模化

数字图书馆要进行有效的信息服务,必须具有一定规模的可利用资源。这种资源可以是网络环境下本地存储的资源,也可以是存储于外部结点的可共享的资源。形成在某个内容或主题上相对独立、系统化的、一定规模的数字化资源是数字图书馆提供服务的基础。只有零散的、不系统的、少量的资源可以利用的信息系统不能称为数字图书馆。

3. 资源的完整性

数字图书馆的目的是直接向读者提供所需的最终信息,而不只是获得信息的线索,即二

次信息,如索引、目录及摘要等。虽然二次信息也可能是某些读者的最终信息需求,书目数据、索引文摘等也是数字图书馆资源的重要组成部分,但是只有索引、目录及摘要等信息线索,而没有最终全文信息的系统不是数字图书馆。

4. 存取的网络化

数字图书馆依附网络而存在,高速的数字通信网络是数字图书馆的存在基础。信息资源的存取将不再受地理位置的限制,对信息资源的使用是远程的。数字图书馆的基本前提是网络环境,没有网络化的存取手段,“图书馆”就不可能完全摆脱传统的图书馆实体。数字图书馆以内部网的形式建立自己的存储机制,构建其内容资源管理框架,最后联入因特网,提供网络环境下的信息存取服务。对内的业务组织和对外的服务都以网络为载体。

5. 管理的分布式

在数字化和网络化的环境下,数字图书馆将以开放和资源共享为基本特征,全球图书馆的模式是数字图书馆管理和服务的方向。在这种模式下,一个数字图书馆不仅要管理和使用自己的馆藏,而且还要利用更大范围的信息资源为读者服务。因而除少数大型数字图书馆外,大多数图书馆将强调信息的存取而不再是馆藏本身,它们将成为信息资源的转接中心、镜像中心、多级体系中的管理中心及信息服务代理中心。少数大型图书馆联合,形成分工合理、经济的信息资源保障体系,成为其他图书馆共同利用的馆藏基础。

6. 服务的透明性

数字图书馆往往是由多个图书馆组成的网络系统,它以网络化存取和分布式管理为特征,因而数据的存取操作可能经过复杂的处理过程,涉及网络环境下不同地理位置的多个结点。但对用户而言,数字图书馆应为各类操作提供简单的接口和使用界面,使用户认为他所面对的是一个完整的数字图书馆整体,其中各种复杂过程对用户是透明的。

六、数字图书馆的特征

(一)文献内容数字化

1. 馆藏文献内容数字化

采用数字技术,将传统图书馆的馆藏,包括印刷型文献、缩微制品和视听资料的内容逐步实现数字转换和处理,并存贮在大容量、高密度的存贮装置中,采用数字图书馆有关数据存贮和标识技术,对数字文本、图像、视频、音频进行分级存贮,用调度系统把它们有机地集成在一起。

2. 电子出版物的加工处理

电子出版物包括电子书、电子杂志、电子报纸等。这些出版物是电子排版文本格式,