●曹学柱 秦莲霞 孟遂珍

国外数字图书馆发展概况

摘 要 国外数字图书馆,从数字化产品的制作到提供网上服务这一过程的全部实用技术已日趋成熟和完善。今后的数字图书馆将与其他一些机构、企业之间在服务内容和方式上越来越趋同;服务向着社会化、国际化方向发展;信息产品的制作将不受部门限制,向着社会化方向发展。参考文献 11。

关键词 国外数字图书馆 技术现状 发展概况
分类号 G250.76

ABSTRACT In foreign countries, practical technologies of digital library have been becoming more and more mature. Services of digital libraries and other institutions will converge in service contents and ways, and will be more socialized and internationalized. The production of information products will not be limited by sectors. 11 refs.

KEY WORDS Digital library. Present status. Development trends.
CLASS NUMBER G250.76

数字图书馆出现至今只有 10 几年,时间虽不长,但发展速度很快,从数字化产品的制作到提供网上服务这一过程的全部实用技术已日趋成熟和完善。及时了解和研究其发展过程,掌握其技术水平,对于加快发展我国的数字图书馆大有裨益。

1 发展历史

国外数字图书馆建设起始于 80 年代中期。1985 年,西班牙文化部,IBM spain 和 Ramón Areces 基金会共同组建的一个联合小组着手于 AGI(EI Archivo General de Indias)数字图书馆建设,目标是建立一个对所有历史档案具有集成功能的计算机系统;该系统的技术核心是收集、描述和管理所有档案及描述信息。西班牙文化部将此项目作为数字图像图书馆的研究起点,故推迟了其他类似项目的研究。1992 年 AGI 正式运行时,所收藏文献(实物)的 10 %的数字图像采集工作业已完成(在光盘上至少汇集了 900 万页的数字图像)。1996 年,在 83000 页的文献复印件中有 38 %来自数字图像而不是原件,保护了由于年代久远或保护不善而已变得异常脆弱的原件,可见该项目在历史文献保护方面的功效。凯斯西部预备大学数字图书馆建设始于 1988 年,侧重于研究知识产权管理以及多媒体全文材料和高质量图像的网络分发,目的是向学生提供更好的教育资源(甚至可以直接送到他们的宿舍)。美国画廊艺术品数字图书馆于 1990 年开发建设,称作 MIRA 项目。因为劣质的图像质量被看作是使用数字图书馆的一个障碍,如何使图像达到最高水准便成为该项目的主要研究问题之一。

90 年代,美国、欧洲一些国家不惜斥巨资开始大规模研制和建设数字图书馆。1993 年美国政府正式提出兴建"信息高速公路"计划,同年提出了一系列由政府出资支持的数字图书馆研究和计划项目。1994 年 9 月,由美国国家科学基金会、国防高级研究项目署、国家宇航局联合发起三机构"数字图书馆倡议",并投资 2400 万美元,于 1997 年底完成。该项目共分 6 个子项目,涵盖大规模文献库、空间影像库、地理图像库、声像资源库。由美国国会图书馆牵头、"美国国家数字图书馆联盟"提出的全国数字图书馆计划,计划在 2000 年前把全国 500 件藏品数字化。由法、日、美、英、德、加、意、俄 8 个国家的国家图书馆参加的"G8 全球信息社会电子图书馆"项目于 2000 年建成,包含各国文化历史精华。日本政府斥资 4 亿美元兴建"日本国会图书馆关西馆工程",预期在 2002 年完成一期工程,该馆将是日本最大的数字图书馆和亚洲地区的文献提供中心。法国"图家图书馆数字化工程",数字资源已达 3000GB 以上,书目数据记录 830 万条。英国"国家图书馆存储创新倡议",共有 20 个项目,大部分已完成,并在因特网上提供服务。俄罗斯"往事项目",包含俄罗斯历史文化精华;俄罗斯"电子图书馆项目",包含 29 个子项目。1997 年,Hemitage 博物馆着手开发博物馆的综合信息服务系统,其核心为数字图像图书馆,以此为依托,提供一个 Web 站点、一个教育中心和一个信息咨询网络。德国政府出资 6000 万马克启动"数字图书馆项目",由德国图书馆、信息中心等机构承担。新加坡"2000 图书馆发展计划",其目的是要使新加坡在华文因特网上占据重要位置。

到 1999 年为止,美、英、法、日四国仅国家图书馆所完成的数字馆藏,其存储量最少约为 500GB,最多的已超过 3000GB。

2　技术现状

近年来,国外以大量人力研究馆藏资源数字化产品制作设备技术、统一界面 Web 化技术、专用的数字图像增强技术、数字图书馆的信息资源的知识产权保护技术、个性化定制和服务系统技术。另外还有许多相关技术,如数据的存储与压缩技术、网络技术、电子商务技术、文字及文字混合处理技术等。它们中的大部分与其他信息系统有一定共性,因此有一些相关领域的成熟技术,拿来稍作开发就可应用。

2.1　数字化产品制作设备技术

数字图书馆建设的一项基础的、大量而又艰巨的工作是把原有馆藏快速完整地转换成数字存放到电子媒体中。特别是对历史文献如手稿、档案、古籍、金石书画、碑拓、简书、帛书等既珍贵又有特色的历史资料处理尤为重要,因此需要根据不同的数字化对象研制不同数字化产品制作设备。梵蒂冈数字图像图书馆和 AGI 数字图像图书馆等项目中,研究人员专门设计了处理手稿原件的架子,既保护原件又提高了速度,每台数字化设备的高质量数字图像生成量日均达到 80 页。

2.2　统一界面 Web 技术

以往的文献数字化产品以光盘为主,使用起来一般要有专门的检索软件,不太方便。光盘数据库上网后,用户还要在本机上安装专用的网络软件,并进行必要的参数设置,才能进行检索,而且各种光盘数据库的检索软件不尽相同,给用户带来许多困难。如果将光盘数据库产品转换可用 Web 浏览器进行检索的形式,就可大大简化界面,节省时间,方便用户。近年来,已经开发出能将光盘产品进行 Web 转换的技术。国外文献数据库开发商纷纷从以光盘产品为主逐步向以 Web 产品为主转变。从发展趋势看,今后的大型数据库将主要通过磁盘陈列和镜像服务器方式向用户提供服务,检索软件将逐步转向基于 Web 的较为统一的界面,并可支持多种语言多媒体的检索。用户只要熟悉 Web 浏览器的使用方法,就能够比较容易获取数字图书馆的各种信息资源,并且大大提高信息获取速率。

2.3　数字图像增强技术

有些文献因为历史久远或保存不善而存在污点、模糊的墨迹或者已被墨迹渗透。为了还其本来面目,科研人员潜心研究设计专门的过滤器。目前有许多此类软件产品可供选择,但也有一些问题需要解决,必须根据数字化对象设计一些专门的数字图像增强技术。

2.4　数字水印技术

数字图书馆建设中遇到的最大问题是知识产权问题,无论是传统图书馆还是数字图书馆,信息安全都至关重要。特别是数字图书馆,由于它的服务大部分是在网上,不如传统图书馆将实物和照片放于馆内那样安全,随时都有被不法分子攫取的可能。世界各地的专家学者进行了多年研究,并提出许多方法,其中数字水印技术是目前常见的知识产权保护措施之一,也是伴随着数字图书馆的发展而产生的特有技术。其特点是在数字图像上明显地或隐蔽地标记数字图像的知识产权,使非授权的商业应用和客户不经申请批准无法使用,有效地限制了数字图像的滥用,梵蒂冈数字图书馆就采用了这种保护措施。

2.5　个性化定制与服务系统技术

个性化定制技术是按用户要求定制特殊用户界面的技术。个性化服务是提高图书馆服务质量的重要措施。数字图书馆可设计专门的计算机系统,对各类用户提供定制的个性化服务。目前已开发出 MyGateway、MyLibrary 等个性化服务系统。在美国华盛顿大学、北卡州立大学和康纳尔大学图书馆中,这些系统已投入使用。这也是数字图书馆发展到一定程度必然出现的且又必需解决的技术之一。

3　发展趋势

数字图书馆在国外已成气候。以数字图书馆为核心的信息资源的开发和利用已经兴起,它将比以往任何时候更加提高收集和使用知识的效率,对以带动大文化为基础的文化经济产业的快速发展,推动社会进步,起着越来越重要的作用。从目前国际上数字图书馆的理论研究、技术水平和 10 多年的实践看,其发展趋向大致有 4 个方面。

(1)为充分发挥数字图书馆能容纳无限规模的、高质量的文献信息资源库群的优势,数字图书馆系统结构设计,从一开始就对网络化提出要求,能否方便上网是系统设计的核心任务之一。如不能与各种网络、通讯系统联接,就不能满足各种条件下的服务和需要,数字图书馆的自身优势就发挥不出来。因此,数字图书馆与网络和各种通讯系统的完美联接,是对它的必然要求。

(2)技术的发展使图书馆与其他一些机构、企业之间在服务内容和方式上具有越来越多的共性,趋同性是数字图书馆发展的趋势之一。数字图书馆信息产品的制作将不受部门限制,向着社会化方向发展,谁有技术、制作设备等条件,谁就可以做。

(3)服务向着社会化、国际化方向发展。各国都在积极研究和完善网上资源管理和利用的具体技术和协议(像知识产权、存取权限、数据安全管理等),以便最大限度地扩大服务范围。

(4)数字图书馆是传统图书馆现代化的必由之路,也是发展方向。我们完全可以相信,随着现代科学技术的高速发展和普遍使用,数字图书馆的技术和作用逐步为人们认识和掌握,研究范围日益扩大,应用领域越来越广。

(下转第 84 页)

是“胃动脉”,通用因素是“功能”,属“事物——方面”型概念限定复合主题。⑤主题概念词的提炼、取舍与转换。其含义是将反映课题内容的、具有实质意义的词汇识别出来,目的是有助于转换成将要使用的检索工具所需要的检索标识。比如“广西近代工业发展概况”,在判断了主题类型和分析出主题结构各因素作用后,会发现其中有实质意义的词为“工业发展史”、“广西”、“近代”,而“概况”无实质意义,可以不用作检索词。若使用《中国分类主题词表》进一步规范主题词,可得“工业经济”、“经济史:近代史”两个主体因素的规范主题词和位置因素词“广西”。只要分析正确,转换成检索语词是比较容易的。能分析出课题中代表中心问题的语词和代表限定部分的语词,并从中删除无实质意义的语词,就基本上掌握了确定主、副主题词的方法。检索入口词的选择不但容易并且准确性也较高。

(5)主题分析方法。对于检索用户来讲,最合适、最容易掌握的主题分析方法是主题结构模式法。就是按照事先设计好的主题结构模式提炼相关主题要素,分析主题要素之间的关系。主题模式可表述为“主体因素——通用因素——空间因素——时间主题——文献类型因素”。课题中包含哪个因素就分析哪个因素,有的因素不只一个就要全部分析出来,以便筛选和进行匹配。比如“90 年代美国政府制定毒品问题的计划”,主体因素是“毒品”、“美国政府”,通用因素是“计划”,空间因素是“美国”,时间因素是“90 年代”。实际检索可以用“毒物走私”、“吸毒”、“美国”、“90 年代”匹配检索文献。需要说明的是,用《中国分类主题词表》规范“毒品”一词,只能用“毒物——走私”、“吸毒”来表示。若有人查找美国政府的各种计划,再使用“美国政府”、“计划”,从隐含角度检索美国政府的计划。因标引人员从隐含角度揭示了这一隐含主题。

使用上述“主题分析”内容设置方式培训用户时,用户能否做出如下分析结果的格式即可检验培训效果。以本文曾使用过的几个课题为例:

(1)“激光在医学上的应用”

主题类型:概念联结型复合主题。

主题结构:中心问题为“激光”(主体因素)、“医学”(主体因素);动态部分为“应用”(通用因素)。

有实质意义的语词:“激光”、“医学”、“应用”。

查表转换:主题词为“激光”、“医学”;副主题词是“治疗应用”(查表得到专指度较高的词)。

(2)“胃动脉的功能”

主题类型:概念限定型复合主题。

主题结构:中心问题是“胃”(主体因素);限定部分是“动脉”(通用因素)、“功能”(通用因素)。

有实质意义的语词:“胃”、“动脉”、“功能”。

查表转换:主题词为“胃动脉”;副主题词是“功能”。

(3)“广西近代工业发展概况”

主题类型:概念限定型主题。

主题结构:中心问题是“工业发展”(主体因素);限定部分为“广西”(位置因素)、“近代”(时间因素)。

有实质意义的语词:“工业发展史”、“广西”、“近代”。

查表转换:主题词为“工业经济”、“经济史:近代史”;副主题词为“广西”。

检索用户如果能够熟练地掌握“主题分析”的上述 5 个内容,那么检索用户与标引人员在对待同一内容的信息资料上将会思路一致。而且能了解各种检索工具或数据库编制者的初衷和编制方法,使用时会感到非常便捷。

参考文献

1　张琪玉. 情报语言学基础(增订第 2 版). 武汉:武汉大学出版社,1997
2　陈界等. 医学信息检索与利用. 北京:中国科学技术出版社,1998
3　李长信等. 医学文献检索. 北京:科学出版社,1995
4　陈树年. 中国分类主题词表标引手册. 北京:北京图书馆出版社,1998

李洪建　河北大学图书馆副研究馆员。通讯地址:保定市合作路 1 号。邮编 071002。
顾潇华　河北大学信息管理系副教授。通讯地址同上。
(来稿时间:2001-07-25)

(上接第 68 页)参考文献

1　周和平. 关于建设中国数字图书馆工程的问题. 中国图书馆学报,2000(5)
2　吴志荣. 论数字图书馆的组织模式. 图书馆杂志,1999(9)
3　余日泰. 数字图像图书馆建设中的主要技术. 图书情报工作,2000(8)
4　李玉安. 电子图书馆、数字图书馆研究与实践述评. 中国图书馆学报,1996(6)
5　王世伟. 论数字图书馆的特点及其对当代图书馆学教育的影响. 图书情报工作,2001(3)
6　赵继海. 数字图书馆发展若干领域的评析. 图书情报工作,2001(3)
7　赵金一. 新千年的挑战:关于中国数字图书馆建设实践的思考. 图书情报工作,2001(3)
8　盛小平. 数字图书馆的知识组织. 图书情报工作,2001(3)
9　党跃臣,张晔,王锰华. 数字图书馆信息资源建设中的版权问题. 图书情报工作,2001(3)
10　范晓鹏. 关于我国数字图书馆建设的探讨. 图书情报工作,2001(3)
11　张菽,李素娥. 数字图书馆建设中的关键问题及其对策. 图书情报工作,2001(3)

曹学柱　秦莲霞　孟遂珍　中国气象局图书馆工作。通讯地址:北京中关村南大街 46 号。邮编 100081。
(来稿时间:2001-06-04)