

哈佛大学图书馆数字化工程介绍

赵 曦

(重庆医科大学 重庆 400016)

摘 要 哈佛大学于1998年启动了一个为期五年的图书馆数字化工程,建设支持数字图书馆馆藏的第一代基础结构框架(技术平台)。本文对哈佛大学就图书馆数字化工程的理解和如何开发一套全面支持数字馆藏(包括分类、检索工具、数据存储、用户界面、检索管理)的核心系统进行了介绍。

关键词 哈佛大学 图书馆数字化 技术平台

哈佛大学有近百座图书馆,地理上分布分散,但是近20年来信息技术及信息共享系统的迅猛发展,使这些图书馆连接起来成为可能。为了避免各个图书馆在数字化方面各自为政,重复建设,造成资源浪费,以及使日益膨胀的数字馆藏风格一致、易于使用 and 把新数字馆藏与哈佛丰富的传统馆藏融合在一起,哈佛大学各院系、图书馆和学校行政部门联合提出了一项为期五年的全校图书馆数字化工程(Library Digital Initiative, LDI),计划耗资1200万美元,打造哈佛第一代数字化图书馆的基础框架,主要用于开展日常工作,并非是一项研究性工程。

建设数字图书馆涉及到各个方面的知识,包括统计数据(包括行政、技术和专业)、数字格式、改革技术和工作流程、授权许可、知识产权、数字对象的存储、用户界面和检索等,在工程开始前和进行中都需要对员工进行这方面的知识培训,同时各个院系还可以通过咨询、项目指南、公共出版物等得到这些知识。

目前,建设数字图书馆所需要的完整的解决方案在市场上不能买到,只有自己开发一套技术基础框架。哈佛大学LDI的技术框架是由三部分组成:a. 馆藏框架,创立、存储、服务于数字馆藏的系统;b. 检索框架,帮助用户检索和理解文献中的内容;c. 普通服务,用框架的许多其它组件组成的一小套核心服务。

1 馆藏框架

1.1 创立 哈佛大学主要通过扫描

和数字相机来进行传统馆藏的数字转化,目前已有两个部门建立了高质量的图像捕捉(通过扫描和数字相机)。这两个部门都提供存储器管理对象所需要的统计数据,并高效率地自动把对象存进存储器中。此外,Kuhn Loeb Music 图书馆已经能非常有效地把声音材料转化为数字格式。

1.2 存储 LDI 框架中最核心的组件之一是存储器,它的作用是对数字文件进行存储、管理、保护等。存储器还保存了两种类型的统计数据:一种是用来管理每一个对象的拥有者、状态、检索规则、对象间关系等的一组很小的数据;另一种是对对象(图像、声音文件等)的个别类型变化的特别统计数据。目前哈佛大学选用 Oracle 数据库作为数据存储,主要用来满足管理者对馆藏的保存管理和风险评估进行分析。

哈佛大学 MIT 数据中心正在开发一套对社会科学数据集特征敏感的功能系统,允许用户检索特殊的数据子集并提供相对简单的探索性分析,以帮助用户决定他所检索的信息是否与他所需要的有关。

1.3 传输服务 一座图书馆的数字馆藏有许多格式,每一个都有它们自己的特殊界面传输应用程序。LDI 馆藏基础框架结构包括6个传输应用程序:a. 单图像传输。这是一个把图像文件从存储器传输到浏览器的简单应用程序。b. 页面跳转器。该应用程序是基于网页的把多页文档转化为单页图像文件的用

户界面应用程序。它是用从存储器调出的XML加密的结构性统计数据文档生成的用户界面,这个界面能提供与特殊部分或文档链接一个文档或一组文档进行描述,文档的每一页都有导航链接(跳到下一页、跳到章、跳到另一卷等)。c. 多媒体。这个程序提供相关的不同类型资源的同步检索(如同步播放和查看声音文件及相关文本文件)。d. GIS(组织地图信息系统)界面。通过浏览器可以对选择的GIS数据库进行检索,并且所有数据库都可以下载到桌面GIS软件中。e. 统计数据界面。社会科学数据能以整体或部分用于统计工具进行分析,对选择性数据库可以制作成简单的表格发布到网上。f. 声音数据流。音乐应用程序正处于开发中,LDI 将准备用 RealAudio 服务器作为音乐数据流的服务。

2 检索框架

2.1 分类数据库(Catalog) LDI 工程非常引人注目的就是提供对馆藏的分类检索。目前哈佛大学正在使用的数据库包括本地 OPAC(又名 HOLLIS),许多文摘、索引数据库和分类检索数据库。已经开始建设6个分类数据库:OASIS 包含加密文档数据库(EAD_s);VIA 人文和社会科学可视材料数据库;BIOL 生物医学图像数据库;Geodesy 哈佛大学组织地图数据库;社会科学数据库,由哈佛大学 MIT 数据中心提供数据;植物标本数据库,由哈佛大学标本中心提供资料。

2.2 跨数据库检索 小数据库比大数据库更容易使用,主题化或格式化数据库更容易使用户找到他们需要的东西。然而数据库多了很容易使用户感到困惑,使用不方便。为了降低数据库使用环境的复杂程度,在哈佛大学 LDI 工程中采用了各种“跨数据库”技术:一种是用一个简单命令就可以对多个数据库进行检索的技术,另一种技术是从多个已有的数据库提取数据的超级数据库技术,很方便地为读者在多个数据库之间进行导航。

2.3 网站 许多图书馆数字化方案都包括建设一些单个网站来对馆藏进行描述和导航。这些网站一般都是手工制作完成,这样不免就会出现各种问题(网站管理者的离去、技术的更新等)。哈佛大学 LDI 方案中将试验一些非常先进的技术:用 XML 加密文档和 XSL 自动生成网站,网站的长期维护将由一些中心程序来完成。

对馆藏资源网站的一个关键要求就是能对非结构性文本文档进行检索(如用未经校对的 OCR 识别出来的文本对扫描的图像文件进行检索)。简单的文本文档可以用单词进行检索,用布尔逻辑式进行检索,相似性检索。

2.4 检索入口 随着电子资源在数量和复杂性两个方面的不断增加,就需要对网站资源进行更好地组织和解释这些东西可以得到。哈佛大学已经开发出了一套功能强大的检索入口系统,它的组织能力更强,能为图书馆用户提供更多的检索入口。

3 普通服务

3.1 检索管理 知识产权和对检索

(上接第 77 页)康发展的。发达国家由于有一个成熟的市场经济体系,社会成员都很重视自己信誉的好坏记录。在利用网络服务时,无论是信息商品的供给方还是求购方所提供的信息都是真实的,可靠的。这就使农业网络的信誉度大大提高,网上交易的成交率大大提高。也就是说,农业网站的服务水平的高低除了

进行限制渗透到数字图书馆的各个角落。检索管理服务通过提供一个单独的机制就可以把许多系统进行共享。检索管理服务有两个重要的功能:a. 个人授权与认证。当前哈佛大学采用图书馆特殊的 PIN,计划下一年就改用一项在全校范围内使用的 PIN 服务。在一种特殊的情况下就可以非常方便地改变密码。b. 团体授权与认证。是以团体为单位进行授权的,而不需要对个体进行认证。

授权认证和 PROFILE 信息存储在用户浏览器的 COOKIES 里,所以用户从一个资源到另一个资源的时候不需要重复提供授权认证信息。不要过于担心泄密,这些 COOKIE 在一个时间限制内就会失效。需要进行认证确认的申请能忽略这些 COOKIE 要求服务器重新进行认证,检索管理服务执行效率特别高。检索管理服务被用于图书馆检索入口系统,允许检索授权的 Internet 资源。

3.2 内容方案 LDI 资金中很重要的一部分(500 万美元)用来支持使数字内容可以通过 LDI 基础框架得到的项目。这些项目可以:帮助设置开发优先权,开发 LDI 框架组件的规划是以项目的需要为基础;通知哪些是真正需要开发的项目;对框架进行测试,一旦框架被开发出来,就需要有真正的申请和真正的用户测试设计的有效性,并提供反馈意见。

在 LDI 的前两年中,已经建立了 9 项工程,内容涉及各个范围的材料:a. 可视资料。把已有的可视材料转化为数字材料是内容转化中最庞大的一项,包括把 19 世纪贸易卡片、亚洲艺术图片、中国历史图片转化为电脑图片。目前,生

物医学图片作为数字转化的样本已经处理完成。b. 文本资料。已经把哈佛大学和 Radcliffe 大学所有年度报告和历史材料都进行了数字化转换。c. 组织信息。组织地形图方案,建立以组织地形图为基础进行检索、数据存储,允许管理者对检索到的数据进行 Web 发布。d. 录音。建立音乐图书馆,该工程涉及到对诗歌、历史悠久的 Duke Ellington 录音和印第安纳经典音乐进行数字转换、保存和网上检索。e. 复合媒体。收录多种类型的数字材料。该工程涉及到对图形、音乐、植物标本、地图等的数字转换和进行注释性的文字描述。

4 目前状态和将来的发展

现在新的检索入口系统、可视材料和文献查找帮助数据库、命名、检索管理服务系统和存储建设等已经完成,所有的内容项目都正在进行,图片和两所大学的数字资料的数字转化快要完成。

单笔投资最大的是建立经过授权的电子期刊网站。哈佛大学将要与其它研究性图书馆合作,建立一个电子期刊检索网站,可以利用 LDI 基础框架结构进行大规模当前电子期刊文献检索。

在以后的日子中,哈佛大学 LDI 最重要的工作是对学校内部资源的处理和把许多不同的外部资源整合到一个相似的系统中,以方便用户使用。

参考文献

- 1 <http://hul.harvard.edu/Ldi/>
- 2 Dale Flecker. Harvard's Library Digital Initiative——Building a First Generation Digital Library Infrastructure. D - Lib Magazine, 2000; (6)

(责编:程钧惠)

网络系统的水平高低、服务水平高低以外,更重要的是与信息商品的交易主体、信息商品的用户的素质高低有着很大的关系。而我国的农业网站服务才刚刚起步就已经出现了网上欺诈行为,这对网络经济的发展是极为不利的。

参考文献

- 1 陶俊昌等. 农业预警系统——农林农业管理的

- 新思路. 北京:中国统计出版社,1992
- 2 郭作玉. 法国的农业信息服务. 世界农业, 2000; (3)
 - 3 张立民, 吴良欢. 新形势下基层农业信息服务的探讨. 1999; (3)
 - 4 唐 研. 美国农产品市场信息收集与发布. 世界农业, 2000; (11)
 - 5 杨有权等. 国外农业生产的咨询服务. 世界农业, 2000; (4)

(责编:亦思京)