

●金 巍 殷立新

非结构化数据库在数字图书馆中的应用

摘 要 iBASE 数据库是非结构化数据库。它具有信息资源数字化建设与收集功能、海量信息资源的存储与管理功能、资源的动态发布与查询访问功能、多元化的信息服务功能和完善的系统管理功能。参考文献 3。

关键词 iBASE 数据库 数字图书馆 信息服务

分类号 G250.74

ABSTRACT iBase is non-structured database and has the functions of digital information resource development, massive information storage and management, dynamic resource distribution and retrieval, diversified information services and perfect system maintenance. 3 refs.

KEY WORDS iBase. Database. Digital library. Information service.

CLASS NUMBER G250.74

非结构化数据库是专门针对关系数据库严格的二维表的结构化而言的,是指数据库的字段长度可变,并且每个字段可以由重复的子字段构成的数据库。iBASE 数据库就是非结构化数据库管理系统软件。下面以 iBASE 数据库为例,阐明非结构化数据库在数字图书馆中的应用。

iBASE 数据库系列产品主要包括 iBASE DMC 数据管理中心、iBASE Webdownload 网上信息采集在内的一系列软件,其功能基本涉及数字信息资源的生产、加工、存储、检索、传递、保护、利用、归档、剔除等全过程。利用 iBASE 数据库所提供的强大功能组件,图书馆人员在专业技术人员的现场指导下,完全可以轻易建立起包括数字化资源建设与收集、数字资源的存储和管理、全文搜索引擎、分布式框架机制和安全信任机制等功能数字图书馆系统体系。

1 信息资源数字化建设与收集功能

馆藏资源的数字化建设是数字图书馆工程的核心,拥有一定规模并从内容或主题上相对独立的数字化信息资源是数字图书馆的基础。iBASE 数据库为用户提供了一个既可以建设自己的数字资源又可以收集其他已经存在数字资源的开放式环境。

利用 iBASE DMC 数据管理中心提供的方便快捷的建库功能,中国矿业大学图书馆建立起了自己特色专业数据库,同时还利用 iBASE DMC 完成了对原有的书目数据库、文摘数据库、全文数据库以及光盘数据倒入 iBASE 数据库中,统一了数据格式,为资源

共享和数据交换奠定了基础。

因特网是最大的信息资源库。为有效利用网上丰富的信息资源,图书馆可利用 iBASE Robot 网上信息采集工具强大的信息收集和自动归类功能,实现信息自动采集、自动分类并直接进入 iBASE 数据库,实现对网络资源的收集和组织管理。

2 海量信息资源的存储与管理功能

存储和管理信息资源的关键是数字化建设的财富如何能被有效地组织、管理,并为用户提供检索服务。利用 iBASE 数据库强大的数据存储与管理功能,可为图书馆提供一个完全基于 B/S 结构的高性能的可扩展的存储和管理信息资源的环境,一个可跨平台的系统。

图书馆可利用 iBASE 数据库作为数据存储和管理的底层结构核心,采用树型目录的管理方式,按照目录-服务器-库-子库-资源的结构,分等级地存储资源,所有数字资源可以按照管理者的意愿,放在他想要放置的地方。同时可利用 iBASE 提供的开放式资源通道,接受和管理所有数字资源,譬如 PDG、PDF、HTML、TXT、DOC、XLS、PPT、XML 以及声音、图像、视频资源和音频资源等;同时利用基于 XML 的开放的数据接口和分布式数据管理能力,实现对跨平台、各种异构数据库的统一数据交换和管理。

对于各种纷繁复杂的海量电子文档资源,iBASE 数据库独特的外挂文件能力为管理这些资源提供强大动力。例如扫描后的 PDG 文件直接作为 iBASE 的

一个字段,利用扫描文件元数据库,通过检索元数据就可直接找到该文件。而各种格式的电子文件可按照内容进行分类,可自动提取标题进行标引,直接入库进行管理。系统还提供远程数据维护管理功能,授权用户可以通过浏览器对数据进行增、删、改等维护管理,提供数据自动备份和恢复功能。

3 资源的动态发布与查询访问功能

将各种资源动态发布上网实现资源共享是数字图书馆的最终目的。全面采用 iBASE 的数据库,并利用其模板组件技术,图书馆无须任何编程就可实现各种文档、数据库信息资源的动态发布和自动更新;同时采用内容和形式相分离的方式,系统完全从数据库中提取数据,按照模板格式自动生成或更新网页,可大大减轻图书馆网站管理人员的负担。

在查询与访问上,图书馆可利用 iBASE 数据库本身具有的强大检索能力,为用户提供查询和访问馆内所有馆藏资源的统一检索入口。用户只需输入一次检索词,就可检索查询馆内的跨多个数据库的信息资源,如 iBASE 数据库、Z39.50 数据库、清华全文镜像数据库等,最后将异构数据库的检索结果统一返回用户;同时利用 iBASE 数据库支持外挂文件功能,实现检索元数据查找全文的功能。

针对不同的资源,系统还提供自动调用资源关联应用程序功能,能被传送和直接在客户端打开,根据资源的格式将自动调用相应的打开工具或者浏览器,比如用户获取一个 PDF 文档,只要工作站有 Adobe 的阅读器就会自动打开它,图片和其他数字资源都是如此。

4 多元化的信息服务功能

基于数字图书馆平台,系统提供基于网络的完善的、多元化信息服务手段。

(1) 信息检索服务:包括公共联机目录检索、光盘数据库检索、自建数据库检索、网上资源库检索、网上资源的组织。

(2) 互动式信息服务:包括网上文献预约、网上参考咨询、网上文献传递、网上文献采购申请、网上馆际互借。

(3) 网上定题服务:包括网上期刊目次通报服务、网上定制服务、网上通报服务、网上文档服务。

(4) 辅助性用户服务:包括图书馆新闻/公告、用户意见箱、BBS、FAQ、网上用户教育等。

5 完善的系统管理功能

根据数字图书馆面临的版权和信息安全问题,iBASE 数据库提供了基于角色的用户权限管理、基于策略的计费管理和完善安全认证管理以及图形界面维护管理机制。

用户权限管理:根据用户角色,提供多级用户管理机制;用户角色不同,相应权限不同,确保只有授权的用户才能访问相关资源。

计费管理:可采用会员制或者按流量、按检索查询资源数量等方式计费,提供完全基于管理策略的计费管理机制。

安全认证管理:提供系统级、数据库级、记录级、字段级四级安全控制,可对系统所有的事件和任务有效监控,提供运行日志、Web 访问等的分析功能;采用因特网的合法 IP 地址,对访问数据类型、端口、时间、内容进行监控、管理,实现网络安全。采用 VPN 技术,保证同客户企业的通信安全。密码、关键数据采用加密传输,通过 HTTPS(SSL 技术)进行,保证客户数据不被篡改,保证网络信息的安全和完整。

维护管理功能:完全基于图形界面,界面友好,操作简单,极大地降低管理员的劳动强度。

数字图书馆的本质就是以分布式海量数据库群为支撑的信息服务系统,利用 iBASE 数据库,图书馆完全可以以统一的标准和规范为基础,进行数字化资源建设并实现对数字文献、网络资源、多媒体等海量信息进行分布式管理,配以智能检索技术为手段,电子商务为管理方式,建立一个功能强大的数字化图书馆信息服务平台,从而实现真正的数字化图书馆建设,为用户提供基于网络的各种信息服务。

参考文献

- 1 北京国信贝斯软件有限公司技术白皮书(1999年9月第1版). <http://www.ibase.com.cn>
- 2 卢共平. 数字图书馆信息的体系结构. 中国图书馆学报, 2002(1)
- 3 张忠友. 非结构化数据库与数字图书馆. 中国信息导报, 2001(8)

金巍 中国远洋运输总公司工程师。通讯地址:北京复兴门内大街158号。邮编100031。

殷立新 中国远洋运输总公司工程师。通讯地址同上。(来稿时间:2002-02-06)