

●谈春梅 田质兵

电子信息资源数据库的开发设计及技术特点*

摘 要 数据库是较为理想的电子信息资源揭示和组织方式。借助于图书馆网络平台采用 C++ Builder 作开发工具,使用新型 XML 技术作检索软件,可开发出电子信息资源数据库系统。系统具有高效、实用等特点。参考文献 5。

关键词 电子信息资源数据库 C++ Builder XML 开发设计

分类号 G250.74

ABSTRACT Database is a comparatively ideal pattern for the description and organization of electronic information resources. We can use library network platform, C++ Builder as the development tool and new XML technologies as the search software to produce a database system of electronic information resources, which have effectiveness, practicality and other features. 5 refs.

KEY WORDS Database of electronic information resources. C++ Builder. XML. Development and design.

CLASS NUMBER G250.74

目前,电子信息资源的揭示与组织主要有主题树和数据库两种方式。一般说来,前者简单、直观;后者信息处理规范,数据格式标准,是一种较为理想的方式。我们借助于图书馆网络平台,采用数据库方式成功地进行了电子信息资源数据库系统(以下简称“系统”)的开发设计。

1 系统开发设计的主要内容与重要功能模块

1.1 系统的主要工作流程

系统的主要工作流程是:数据库电子信息资源、因特网电子信息资源→电子信息资源 MARC 模板设计→MARC 数据录入→MARC 数据通讯格式→MARC 数据库↔SQL Server 2000 关系型数据库→数据库网络查询→查询结果输出。

从工作流程可以看出,系统开发设计的内容坚持以人为本,紧紧围绕用户的需求,对有价值的数据库电子信息资源和因特网上的电子信息资源进行有效搜集,采用先进的 C++ Builder 开发工具,使用面向对象开发方法编程,进行电子信息资源 MARC 模板设计,生成 MARC 数据通讯格式和 MARC 数据库。同时,经过相关程序的开发设计,将 MARC 数据库转换为 SQL Server 2000 关系型数据库系统。并且

摒弃传统的 CGI(Common Gateway Interface,公共网关接口)、ASP(Active Server Pages,活动服务器网页)数据库查询编程方法,选用新颖的 XML(Extensible Markup Language,可扩展标注语言)技术进行检索软件的设计,以完全基于网络应用的新的高质量的电子信息资源数据库,供用户检索和利用。

1.2 系统的重要功能模块

系统分为两大功能模块,即电子信息资源 MARC 数据库管理与维护模块和网络查询模块。每个主模块中又有若干个功能模块。

MARC 数据库管理与维护功能模块主要有:电子信息资源模板设计、MARC 编目、MARC 数据库维护和代码维护等。每个功能模块下又有若干个子功能模块。

电子信息资源模板设计主要包括:中文电子期刊模板设计、外文电子期刊模板设计和外文电子会议录模板设计等子功能模块;电子资源 MARC 编目主要包括:加载模板、增加字段、生成 MARC、重要字段提取和相关代码等子功能模块;电子信息资源 MARC 数据库维护主要包括:查询和修改等子功能模块;代码维护主要包括:中外文电子期刊地区代码和出版频率等子功能模块。

* 本文系江苏省哲学社会科学规划基金项目(2001 省 05)的研究成果之一。

网络查询功能模块主要有:表单提问、网络基本查询以及高级查询和查询结果数据显示等。

各个功能模块既相互独立又相辅相成。这些功能模块已实现系统设计的技术要求并能正常运行。

2 系统的主要技术特点

2.1 坚持数据格式标准化原则

数据格式标准化是电子信息资源数据库系统的基础,也是网络应用和发展的关键。系统主要选择通信技术、计算机技术、经济管理、教育等重点学科的中外文电子期刊、外文电子会议录等,并采用了国际公认的、比较成熟的 MARC 数据格式。中文电子期刊采用中国机读目录格式 CNMARC 标准数据结构,外文电子期刊和外文电子会议录采用 USMARC 标准数据结构,并建立合理的简要式 CNMARC、USMARC 标准数据结构体系。中文电子期刊设有 39 个字段,外文电子期刊设有 30 个字段,外文电子会议录设有 23 个字段。上述所设字段不但比较简洁,而且覆盖了“都柏林核心”15 个字段的全部内容,能够充分完整地记录和描述电子信息资源。

2.2 每条记录均进行主题标引

主题标引是衡量电子信息资源 MARC 数据库质量高低的一个重要标志。我们以《中国分类主题词表》、《邮电通信主题词表》和《中国教育分类主题词表》作为标引工具,对于每条中外文电子期刊和电子会议录均进行了主题标引,有效地提高了广大用户电子信息资源的查准率。

2.3 采用先进的 C++ Builder 进行前台数据库开发

C++ Builder 是基于 Windows 的面向对象的可视化开发工具,具有强大的数据库处理能力,全面支持基于组件的应用程序开发。与以往人们常用的 VB (Visual Basic)、PB (Power Builder) 等开发工具相比, C++ Builder 开发工具有着许多优点。我们在程序编写和调试过程中,有效地利用了 C++ Builder 提供的 ADOTable 和 ADOQuery 等组件,实现了 MARC 字段数据超过 255 个字符的录入、服务器时间的自动提取(记录板本标识,005 字段)、出版地区代码和出版周期代码的自动生成等功能。进而实现了电子信息资源 MARC 数据通讯格式的快速生成以及 MARC 数据库与 SQL Server 2000 数据库的快速转换。MARC 数据库可用于网上数据交换,实现电子信息资源数据共享。

2.4 以新型的 SQL Server 2000 作为电子信息资源

后台数据库

SQL Server 2000 具有高度可伸缩性和可靠性,尤其是能够提供 XML 和 HTTP 的全方位支持,为电子信息资源数据库 Web 应用提供了完善的数据管理和数据分析解决方案。我们正是利用 SQL Server 2000 这些特点进行 XML 数据库网络查询程序的开发和数据管理,使数据库安全、可靠、易操作。

2.5 使用新颖的 XML 技术设计检索软件

我们摒弃了传统的 CGI、ASP 查询编程方法,采用了新的组件与 XML 综合开发方式,设计新型的检索软件。在使用 XML 技术开发检索软件过程中,突破了乱码输出、模糊检索等多个难点问题,较好地实现了电子信息资源数据库 XML 模板基本查询以及 XML 的高级查询。XML 高级查询是对客户端浏览器上以模板查询方法得到的查询结果进行直接检索,不再经过服务器,使原先集中于服务器端的大量运算负荷分布到客户端,从而大大降低了服务器的消耗,缩短了响应时间,提高了网络效率。

2.6 调用函数(indexof, substring)处理查询结果

如何在查询结果中输出多个超链接网址和与之对应的公用附注、专指资料等信息,是本系统开发设计的重要任务。在电子信息资源 MARC 编目中,我们设置了“856”字段及其子字段@u(统一资源定位地址,即 URL 地址)、@z(公用附注,提供全文、题录等需要说明的内容)、@3(专指资料,说明数据库所收录的电子信息资源的年代范围等信息),其目的就是通过该字段的提供,实现对数据库的超链接。

由于在 XML 技术中,字符串处理功能比较薄弱,我们使用 XSLT-XPath 结合 Java-Script 语言编程。在规定好不同超链接网址、不同公用附注、不同专指资料之间用标识字符“;”来分割的前提下,又定义了超链接网址和与之对应的公用附注、专指资料等信息为一个“节点”。我们首先读取这个“节点”的内容,然后调用函数(indexof, substring)来判断、截取这个节点中的信息,并用 CreateNode 建立与超链接网址数目相等的子节点,再为这些子节点建立几个新的子节点,将截取出来的信息存储在新子节点之中。

在格式化输出的时候,我们在 XSL 里面规定输出的是这些新子节点的内容。这样就实现了在查询结果中输出多个超链接网址,并把超链接网址和公用附注及专指资料对应起来的功能。这样能够使用

户在同一检索平台上检索多个数据库和多个网络电子信息资源。

3 系统应用

经过一年多的开发和研制,系统已经由开发研制阶段转化为实际应用阶段。由于系统采用了先进的 C++ Builder 开发技术和新颖的 XML 查询编程方法,较好地解决了 MARC 数据通讯格式的生成、MARC 数据库到关系型数据库的相互转换等关键问题,有效地突破了 XML 程序设计中的高级查询、查询结果输出等难点,使得系统运行效率高、实用性强,安全可靠。

3.1 电子信息资源 MARC 数据库的建设

电子信息资源 MARC 数据库建设是网络查询的基础和关键。建库工作者按照系统所提供的功能进行实际操作,根据系统所提供的中外文电子期刊、外文电子会议录等简要式 MARC 模板,对于具有权威性和实用性的数据库,如万方网络数据库、ACM 数据库以及 IEEE 出版物等数据库的全文电子期刊、电子会议录和因特网上的电子期刊、电子会议录等信息资源,进行 MARC 字段数据的填写与录入。由于 001 记录控制号、005 记录版本标识、008 和 100 通用处理数据等字段内容是由系统自动生成的,出版地区代码和出版周期代码又可由建库工作者根据实际需要自动选择填写,这就给建库工作者带来了极大的方便。此外,中外文电子期刊,605、606(主题标引),610、653(非控主题词),710、711(主办者团体名称),856(统一资源定位地址)等 MARC 字段都可以重复地录入与存储,从而大大增加了检索点。尤其是 856 字段的重复,为在同一检索平台上检索多个数据库以及多个网络电子信息资源提供了基础和条件。实践证明,电子信息资源 MARC 数据库建设既方便又快捷。

3.2 网络查询

系统上网运行后,用户可通过网络直接访问南京邮电学院图书馆电子信息资源数据库服务器,使用 XML 基本检索和高级检索两种方式进行多途径检索。

系统主要提供题名、缩略题名、主办者、出版者、ISSN、分类号、主题词、关键词、内容描述以及组配等检索条件。用户检索的基本步骤为:检索途径的选择和检索内容的填写、基本检索、高级检索和超链接。

例如,有一位搞科研的博士生,急需有关通信(communications)方面的外文期刊。他进入系统后,按照系统的提示进行检索。在现有的数据库中共检索到了 12 条相关记录。他又进行了高级查询,检索其中的一条记录,如 *Communication World*,系统将该条记录的详细信息显示如下:

电子资源标识:Electronic Periodicals

出版频率:Monthly

分类号:TN

ISSN:0744-7612

电子资源定位地址:

<http://global.umi.com/pqdweb?RQ>

[T=317&](#)

[sk=1&SCQ=?&Tag=1&INT=0&S](#)

[ellLanguage=0&TS=1038383391](#)

公用附注:Full Text

专指资料:ProQuest Full text Coverage: Dec 1987
(Volume4, Issue12) current

他点击上述“电子资源定位地址”进行超链接后,快速进入 UMI 数据库,查阅到了 *Communication World* 的全文。

电子信息资源数据库系统开发设计的成功,不仅使图书情报界信息服务领域电子信息资源实现规范化、标准化的存储,而且为电子信息资源网上数据交换和查询提供了软件支持,从而满足广大用户对电子信息资源集成化、高效化和社会化服务的需求,具有极其重要的推广和应用价值。

参考文献

- 1 肖延松等.基于 XML 的数据交换在分布式 web 系统中的应用.电脑与信息技术,2002,10(1)
- 2 谈春梅等.电子信息资源数据库检索系统的开发与实现.中国图书馆学报,2002(6)
- 3 <http://www.2tigers.net>
- 4 飞思科技研发中心. SQL Server 2000XML 应用开发.北京:电子工业出版社,2001
- 5 杨宗志. C++ Builder 数据库程序设计.北京:清华大学出版社,2001

谈春梅 南京邮电学院图书馆副研究馆员。通讯地址:南京市。邮编 210003。

田质兵 南京邮电学院图书馆助理馆员。通讯地址同上。
(来稿时间:2003-05-19)