

数字资源跨平台整合系统的设计与实现

周育红¹, 闫锋欣²

(1. 陕西科技大学 图书馆, 陕西 西安 710021; 2. 西北工业大学 机电学院, 陕西 西安 710072)

摘要: 为进一步提高图书馆大量数字资源的整体使用效率, 提高数字资源和馆藏文献的检准率和检全率, 解决现有馆藏检索系统难于跨平台运行的问题, 提出了一种新的图书馆数字资源参数模型: 在研究都柏林核心元数据 DC 字段和常用的 MARC 数据格式之间的映射关系基础上, 给出了数据库概念模型的设计方法, 实现了整合图书馆馆藏文献和本地其他信息资源的跨平台原型系统 DRMS, 并从开发目标、框架设计、开发平台等方面概述了系统的设计思想。最后, 以 C++ 和 Qt 应用程序框架作为系统开发语言, 完成了 DRMS 的数据管理模型设计、软件开发和系统功能, 为用户提供了多平台操作系统下的统一人机交互界面和检索方法。

关键词: 数字资源; 跨平台; 整合系统; Qt 应用程序框架; 设计与实现

中图分类号: TP391

文献标识码: A

文章编号: 1673-629X(2010)01-0243-04

Design and Implementation of a Cross-platform Digital Resource Integrated System

ZHOU Yu-hong¹, YAN Feng-xin²

(1. Library of Shaanxi University of Science and Technology, Xi'an 710021, China;

2. Mechatronic School of Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: In order to improve the overall usage efficiency of the current digital resources, enhancing the search accuracy rate of literatures existed in libraries, resolving the problem that current research systems cannot run well in different operating systems, a novel library information resources integration model is presented. On the basement of studying of the mapping relationship between Dublin Core fields and MARC fields, a cross-platform prototype system which integrates library collections and other local information resources is designed and implemented. From the development goals, framework design, development platform, provides an overview of the design ideas. Finally, based on C++ and Qt application framework, a system was developed and some experiments shown in the last sections validated the good effects.

Key words: digital resources; cross-platform; integrated system; Qt application framework; design and implementation

0 引言

近年来, 高校图书馆通过集团购买、自主采购和自建数据库等方式完成了数字图书馆建设。这些引进或者自建的高质量数字资源信息数据库, 显著增加了以电子图书、期刊和联网数据库为代表的数字化信息资源, 为读者提供了更多的信息来源和获取途径; 但不同数据提供商的数字信息资源对数据库运行的软硬件系统、环境各不相同, 每个数据库支持的检索运算符和检

索语言也不尽相同, 造成了数字资源信息检索的差异性和复杂性^[1]。因此, 为有效满足面向图书馆的复杂应用对大规模计算能力、海量数据处理和信息服务的需求, 将广域分布的异构、自治资源进行按需组织和管理, 以实现动态、跨自治域的资源共享与协同, 提高信息资源的综合利用率已成为一个重要的科学问题^[1~3]。

数字资源检索模式和资源共享与集成运行环境的统一能够有效解决上述问题, 目前, 国内外学术界和工业界已开展了大量的研发工作。张宇娥^[1]研究了数字图书馆建设中数字资源的整合工作的重要性; 徐汝兴^[4]提出一种跨平台检索系统构建模式, 通过实践经验对其做了初步完善, 但并不是真正的跨平台系统; 姜爱蓉等^[5]认为资源整合与门户建设是现阶段图书馆提高资源利用效率、提升服务应对能力的重要标志; 段智

收稿日期: 2009-04-13; 修回日期: 2009-07-29

基金项目: 国家 863 高技术研究发展计划资助项目 (2006AA04Z111); 国家自然科学基金资助项目 (50475145)

作者简介: 周育红 (1969-), 女, 重庆人, 副研究馆员, 研究方向为信息检索与情报分析; 闫锋欣, 助教, 博士研究生, 研究方向为计算机辅助技术。

敏等^[6]基于校园网和一卡通提出了“1+X”的资源整合模式,但其仅能用于 Windows 系统;Wojtczyk 等^[7]给出了建立跨平台应用程序的 C++ 工作流程。这些研究为建立数字资源统一规范的采集、检索管理体系、实现资源共享提供了必要的理论探索和技术支撑。

1 系统设计

开发能够支持中英文的国际化用户界面和操作环境,无缝运行于 Windows 和 Linux 等操作系统间的跨平台数字资源管理系统(Digital Resources Management System, DRMS)已成为图书馆未来创新服务的支撑点之一。

1.1 开发目标

DRMS 以图书馆资源管理的需求为目标,实现图书馆文献和电子资源参数的规范化管理,提供电子资源的数据分析功能,构建数字资源文献管理应用和研究的基础数据平台,提升图书馆服务职能和工作效率。同时,考虑到实际应用中图书资料等文献信息的特殊性和图书管理人员知识结构的差异性,要求系统运行稳定、可靠,操作简便、实用,系统扩容性、通用性要好,以适应多样的运行环境。

1.2 系统框架

DRMS 由电子资源数据库、前端用户交互界面和数据应用接口函数库三部分组成。在程序设计和系统集成方面,采用分层次、组件式、模块化方式,如图 1 所示。

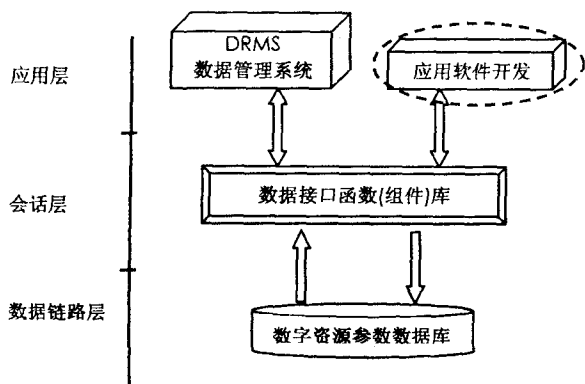


图 1 DRMS 系统框架设计

数字资源参数数据库是数据组织、管理和存储的载体。对数字资源的内容等信息进行分类、归纳和抽象,建立相应的概念数据模型;以数据库主目录为根、书目(期刊)为对象、分层次引伸若干数据子集,构建关系型数据管理模型。

数据库前端用户交互界面是操作数据库和数据应用的平台。通过可视化的人机交互方式,实现数据的管理、浏览、查询、分析、显示等各种应用功能,完成任

务处理。数据接口函数(组件)库是向应用软件开发提供数据支持的桥梁,该部分独立于数据库系统之外,通过接口函数实现对数据库中包含数据的全面操作。

1.3 开发语言

对现有的跨平台技术分析和对比基础上,选择 C++ 语言和 Qt 应用程序框架^[8]作为系统的开发语言,数据管理采用自定义关系型数据结构,主要考虑的因素如下:

- * C++ 是最为流行的程序开发语言,不依赖于任何虚拟运行环境;

- * Qt 自身是 C++ 类库,具有良好的程序代码重用性,程序可运行于 Windows/Linux/Mac OS/嵌入式平台等;

- * 系统不需其他商业软件支持,具有完全的自主知识产权,可方便用于后期的商业推广和应用;

在 Windows 中,集成开发环境(IDE)通常是 Visual Studio 系列^[9]或 C++ Builder^[10,11],可管理项目的程序开发工作。在基于 X11 的 Unix/Linux 系统中,较成熟的 IDE 有 KDevelop 和 SlickEdit 等。KDevelop 可提供类似于微软 VS 的开发环境,功能相对完善,是 Unix/Linux 平台下常用的标准 IDE;SlickEdit 是一款跨平台的 C# 开发工具,但需通过配置来提供对 Qt 的支持。因此,在 Windows 平台下选用 VS 2005、Unix/Linux 平台下以 KDevelop 作为系统的集成开发环境。

2 数据库概念模型设计

数据库概念模型的设计是对目标包含信息的收集、综合分析、描述、归纳和抽象,建立数据分类和层次逻辑关系,以数据安全、操作方便和高效为目标。对 DRMS 系统来说,数据库概念模型设计的基础就是数字资源参数模型。

2.1 数字资源参数模型

对图书馆中数字资源文献所涵盖内容的采集、检索和管理,一般通过都柏林核心元数据的 DC(Dublin Core)格式进行表示,如题名(Title)、权限(Right)、语种(Language)、类型(Type)、格式(Format)等 15 个字段,但实际应用中通常难于同时满足 DC 字段中的全部内容。对此,在对数字资源参数的组成、特征和来源进行研究的基础上,结合常用的 MARC(或 CNMARC)元数据格式,文中构建了一种新的数字资源参数分类模型,如图 2 所示,以充分揭示数字资源的内容和其他特性。

2.2 MARC 与 DC 的转换

DC 的 15 个基本元素分为三部分:内容描述、外部属性描述和资源责任描述,主要是对文件在形成和保管时信息的自动抽取。MARC 格式也包括三部分:头

标区、目次区和数据区,是将图书馆的各种信息资源进行综合分析处理,提取主要信息要素并按一定规范模式组织起来的信息段。两者各有侧重,无法相互完全取代,但在主要内容上两者一致,可以互相转换。

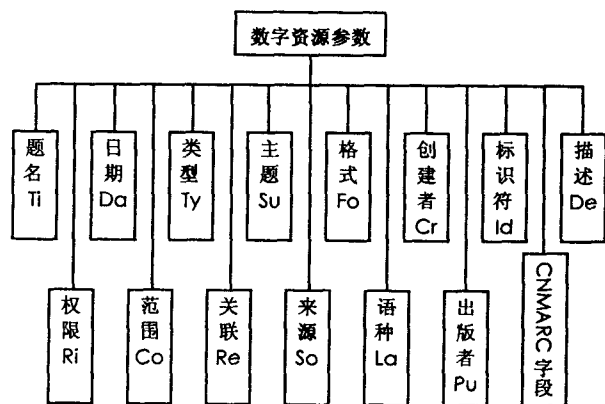


图 2 数字资源参数模型

要实现 MARC 与 DC 字段的相互转换,必须先实现 MARC 与 DC 元素之间的映射^[12]。DC 与 CNMARC 之间的映射关系如表 1 所示。利用这一映射关系,可充分利用网络资源创作者或提供者编制的 DC 数据,将其中的部分数据转换为 MARC,为 DRMS 系统提供丰富的数据资源。

表 1 DC 字段与 CNMARC 之间的映射关系

DC	CNMARC	DC	CNMARC
So	324	Re	332
Pu	210	Ty	215
Co	307	Cr	905
Su	690,692	Da	210
Fo	336,337	La	101
Ri	021	De	300,330,327,600,601,605,606
Ti	200,510,517,540	Id	001,010,011,091
Co	200,701,702,711,712		

3 数据接口函数库的开发

数字资源参数在图书文献管理中的应用不仅是简单的数据统计分析,多用于用户检索、藏书描述、新进文献预测等环节,与相关学科在方法、技术、信息等方面相关。因此,DRMS 系统还要为信息的综合应用和深层次的技术开发提供便利、实用的数据平台和环境。

3.1 接口函数的分类

数据接口函数是数据库连接应用程序的桥梁。DRMS 系统采用组件式、模块化的集成设计,数据接口函数与系统相独立,可延伸数据库的应用,满足进一步研究开发的需要。接口函数可分为 5 种:1)图书馆数据;2)书库数据;3)数字文献标识参数;4)非数字文献类别参数;5)操作参数。此时,以图书馆名称、书库名称、文献名称、检索编码和特征参数构建层次关系,实

现接口函数的阶梯式设计,满足不同层面的数据需求。

3.2 数据查询

通过上述设计所得的接口函数,可以实现资源的动态分配、释放,以及对数据库各类数据进行读写、查询、增加、修改和删除操作。数据接口以函数库的形式提供给用户。表 2 给出了用于接口函数查询的数据表样例。

表 2 接口函数查询

No.	Ti.	Au.	So.	Da.	La.	Fo.	...
0	浅析中文图书 3XX 字段存在的问题	余望枝	职业时空	2008/09	CN	EB	...
1	简述计算机技术在企事业单位档案工作中的应用	王雪瑞	档案	2008/06	CN	EB	...
2	太原市城建档案馆档案数字化建设	尹喆	城建档案	2008/09	CN	EB	...
3	高校图书馆业务外包关系中的冲突分析和质量控制	李先业	农业图书情报学刊	2008/10	CN	JN	...
4	Interaction Techniques and Vessel Analysis	Bernhard	MICCAI 2000	2000/10	EN	EB	...
5	西文回溯编目常见问题分析及对策	谢燕	农业图书情报学刊	2008/10	CN	EB	...
...

4 系统界面功能概述

DRMS 是一款可视化的跨平台数字资源应用系统,可用于完成数字资源参数数据库的创建、维护、使用和管理功能,包括 8 个模块:用户管理、数据采集管理、资源管理、DC 维护、CNMARC 维护、资源查询、资源分析和系统帮助。

4.1 用户管理

系统采取用户注册的方式进行用户管理,可通过超级用户进行用户注册和授权,登录后即可进入相应权限的数据库处理模块。用户管理模块包括用户注册、权限设置、删除用户、口令设置修改等。用户权限分为管理级用户、一般用户和浏览用户。

4.2 数据库管理

数据库管理包括数据库的创建环境设置、资源分配、口令设置修改、数据库备份、删除数据库、退出系统。为了安全和保密起见,前项操作是数据库管理员的权限,数据库口令则仅由数据库管理员掌握。

4.3 资源管理

系统以文献为对象进行数据分类管理(如表 1 所示)。文献管理功能包括建立文献索引、设置修改文献索引口令、删除文献索引、文献索引列表和选择。

4.4 DC 维护

数字资源文献数据可以通过 DC 字段或者 MARC (或者 CNMARC)形式进行管理。此时,可以通过数字资源文献的 DC 或 MARC 字段对其进行检索、增删、输出等操作。用户也可以调用系统中的子功能,如信息转换功能,根据所得文献的 DC(CNMARC)字段域值

对其进行快速转换,得到相应的 CNMARC(DC)字段的域值,从而提高资源的共享效率。

4.5 资源查询

数据查询是 DRMS 系统前端用户界面的主要功能之一,系统提供交互式查询工具,可进行多条件组合和模糊查询,也可自行设计查询数据表的显示格式,如将“语种”的 CN 字符显示为“中文”,“类型”JN 显示为“过刊”。查询结果可通过多页数据表的形式进行分类,与之关联的数据则可输出显示成文本形式。

4.6 数据分析

数据分析和数据统计是图书馆日常管理的有效方法,也是用户对其所关注数据进行处理的方法之一。对 DRMS 的实现方法而言,如果直接将结果输出,一般用户难于理解其真实含义,使得结果分析较为困难。因此,在对检索结果进行处理的同时,还可以对其进行数理统计分析,如统计和预估出读者对于何种文献在特定时间内的需求量,总结不同文献在特定关注条件下所表现出的变化规律和参数特征,为图书馆的文献订阅和服务模式的改进提供直观依据。

5 结束语

数字资源管理系统 DRMS 灵活、方便、实用,为图书馆学应用于信息检索构建了良好的数据分析和数据应用平台。DRMS 系统的数字资源参数数据库、数据库前端用户界面和数据应用接口函数库的设计,方便了不同层面的应用需求面向对象的组件化软件设计,便于数据管理扩容和功能扩充。由于系统不需要商业软件的支持,对硬件环境没有特别的要求,因此,具有较强的灵活性和可移植性,既可以单独运行,也可以作为软件包组装到其它软件系统上使用,取得了良好的应用效果。

(上接第242页)

性,以达到提高政府办公效率、促进经济发展的目的。而数据仓库、数据挖掘正是实现政府决策支持的核心技术。以数据仓库、数据挖掘为依托的政府决策支持系统,将发挥重要的作用。

参考文献:

- [1] 吉同路. 政府资源计划(GRP)初探[J]. 哈尔滨工业大学学报, 2003, 35(s): 132 - 136.
- [2] 王 珊. 数据仓库技术与联机分析处理[M]. 北京: 科学出版社, 1998.
- [3] Inmon W H, Hackathorn R D. Using the Data Warehouse [M]. New York: A Wiley Q-ED Publication, 1994.

参考文献:

- [1] 张宇娥. 数字图书馆建设中数字资源整合研究[J]. 电子科技大学学报, 2002, 31(s): 42 - 44.
- [2] William H, Sameer A, Rodney L L, et al. SPIRS: A Web-based image retrieval system for large biomedical databases [J]. International Journal of Medical Informatics, 2009, 78 (4): 13 - 24.
- [3] Wendy O, Daljit K, Kathy C, et al. A cross-platform solution for bibliographic record manipulation in digital libraries [C]//Proceedings of the Sixth IASTED International Conference on Communications, Internet, and Info. Technology. July 2 - 4, 2007. Banff, AB, Canada: [s. n.], 2007: 193 - 198.
- [4] 徐汝兴. 图书馆跨平台信息检索系统初探[J]. 上海交通大学学报, 2003, 37(s1): 191 - 194.
- [5] 姜爱蓉, 黄美君, 姜天芳. 数字资源整合与信息门户建设[J]. 现代图书情报技术, 2006(11): 2 - 6.
- [6] 段智敏, 王如龙, 孙美青. 基于一卡通的数字化校园资源整合研究与实现[J]. 计算机工程与科学, 2008, 30(1): 8 - 11.
- [7] Martin W, Alois K. A cross platform development workflow for C/C++ applications [C]//The 3rd International Conference on Software Engineering Advances, Sliema, Malta: [s. n.], 2008: 224 - 229.
- [8] Blanchette J, Mark S. C++ GUI Programming with Qt 4 [M]. 2nd Ed. [s. l.]: Prentice Hall, 2008.
- [9] Powers L, Snell M. Visual Studio 2005 技术大全[M]. 刘彦博, 肖 鹏, 贾 蕊, 译. 北京: 人民邮电出版社, 2008.
- [10] 王 晨. C++ Builder 数据库开发经典案例解析[M]. 北京: 清华大学出版社, 2005.
- [11] Hollingworth J, Swart B, Cashman M, et al. Borland C++ Builder 6 Developer's Guide [M]. USA: Sams, 2002.
- [12] 罗庭芝. 从编目角度探讨图书馆数字资源的建设[J]. 当代图书馆, 2008(1): 48 - 51.
- [4] Han J W, Micheline K. Data Mining Concepts a Technique [M]. Beijing: High Education Press, 2001.
- [5] Biswas G, Weinberg J B, Fisher D H. Iterate: A Conceptual Clustering Algorithm for Data Mining [J]. IEEE Transactions on Systems, Man, and Cybernetics (Part C), 1998, 28(2): 100 - 111.
- [6] Chen M S, Han J, Yu P S. Data Mining: An Overview from a Database Perspective [J]. IEEE Transactions on Knowledge and Data Engineering, 1996, 8(6): 866 - 883.
- [7] Glymour C, Madigan D, Pregibon D, et al. Statistical Themes and Lessons for Data Mining [J]. Data Mining and Knowledge Discovery, 1997, 1(1): 11 - 28.
- [8] 蒋海琴, 阚国年, 蒋文明, 等. 房产管理信息系统 [M]. 北京: 科学出版社, 2007.