

# 基于开源少民信息资源保存系统设计研究

龙飞<sup>1</sup>, 于蕾<sup>2</sup>, 杜昌钰<sup>1</sup>

(1. 贵州民族学院 计算机与信息科学学院, 贵州 贵阳 550025;

2. 贵州民族学院 人文科技学院 计算机科学系, 贵州 贵阳 550025)

**摘要:**互联网和信息处理技术飞速发展的今天,如何收集整理、保护和利用少数民族的信息资源,传承中华民族文化,增进民族之间相互尊重、相互团结和相互交流,成为了研究课题。本着这个目的,针对我国各少数民族信息资源保存的现状存在的问题,提出在遵循 OAIS 参考模型的前提下,利用开放源代码数字资源管理系统 DSpace 设计少数民族信息资源长期保存及开发系统架构。此方案目前已运用在少数民族信息资源决策系统的平台搭建上,实践证明它是可行的。

**关键词:**少数民族;信息资源管理;功能模型;数字仓储

中图分类号:TP39

文献标识码:A

文章编号:1673-629X(2011)12-0227-03

## Minority Information Resources System Design and Research Based on Open Source

LONG Fei<sup>1</sup>, YU Lei<sup>2</sup>, DU Chang-yu<sup>1</sup>

(1. School of Computer and Information Engineering, Guizhou University for Nationalities, Guiyang 550025, China;

2. Department of Computer Science, College of Humanities & Sciences of Guizhou University  
for Nationalities, Guiyang 550025, China)

**Abstract:** With the rapid development of information processing technology and internet, how to compile, protect and use information resources of minorities, the Chinese cultural heritage, promote mutual respect between peoples, solidarity and mutual exchange, has become a research topic. This information resources for preservation of the status of ethnic minority problems, follow the OAIS reference model proposed in the premise, the use of open source DSpace digital resource management system designed long-term preservation of information resources and ethnic minority development system architecture. Practice has proved that this is feasible.

**Key words:** national minority; information resources management; functional model; digital storage

## 0 引言

中国有 56 个少数民族,每一个民族在不同的历史时期曾创造本民族的文化,并保存了大量的文献记录,这些丰富的民族文献信息资源将有利于民族文化的传承和发扬、有利于促进少数民族地区经济的发展、有利于深入的学术研究。本课题旨在建立一个较为全面完善的少数民族数字资源信息系统以实现信息资源的长期保存和持续存取,为少数民族地区的经济建设和文化发展服务<sup>[1]</sup>。

## 1 OAIS 参考模型

OAIS(Open Archival Information System)开放档案

收稿日期:2011-04-14;修回日期:2011-07-21

基金项目:国家民委 2009-2011 科研课题(09GE04);贵州民族学院 2009-2010 校级科研课题(校科研 2009JS06)

作者简介:龙飞(1978-),女,讲师,主要从事计算机应用;杜昌钰,副教授,主要从事网络及信息安全。

信息系统的简称,是一种开放性数字信息系统建设的基本框架和所应遵循的原则,由美国空间数据系统咨询委员会(CCSDS)制定的标准,于 2003 年作为 ISO 的标准颁发。OAIS 的功能包括摄入功能(Ingest)、档案存储功能(Archival Storage)、数据管理功能(Data Management)、系统管理功能(Administration)、保存规划功能(Preservation Planning)和存取功能(Access)<sup>[2,3]</sup>。

OAIS 参考模型将与数字文档系统相关的环境、系统的功能模块、信息元数据进行概念化,它为少数民族信息资源系统提供较好的参考模型。

文中遵循此参考模型,结合数据库设计、系统的功能结构、存储过程、处理平台等技术方面,提出少数民族信息资源系统架构。

## 2 开源软件平台

目前,国际上开发的数字存储开源系统有 Eprints(由英国 Univ. of Southampton 开发)、Fedora(由 Univ. of

Virginia and Cornell 开发)、GreenStone (由 Univ. of Waikato 开发)、DSpace (由 MIT 和 HP 联合开发)。其中,DSpace 以其高度的可扩展性、兼容性以及界面友好、发展前景看好等特点,受各大高校及研究机构的青睐,成为数字存储开源系统中用户最多的一个。少数民族信息资源保存及管理系统也将选择 DSpace 作为底层仓储系统。

DSpace (数字空间) 系统是由美国麻省理工学院图书馆 (MIT Libraries) 和惠普公司实验室 (Hewlett-Packard Labs) 合作研究开发的,于 2002 年 10 月开始投入使用的,以内容管理发布为设计目标,面向研究机构使用的开放源代码数字存储系统。该系统可以收集、存储、索引、保存和重新发布任何数字格式、层次结构的永久标志符研究数据。

DSpace 系统分为存储层、业务逻辑层和应用层三个层次,每一层都由一系列功能组件组成。存储层主要负责元数据及内容方面的物理存储,业务逻辑层由一系列功能组件组成,包括检索、浏览、内容管理、电子用户管理、权限管理、 workflow 管理等。应用层主要负责 DSpace 与外部终端用户的交互<sup>[4-9]</sup>。

### 3 系统功能分析及确立

#### 3.1 系统分析

文中先对少数民族信息资源系统的目的与需求,资料搜集的范围和内容,资料提交和提供使用的对象,以及如何长期保存、管理与提供服务做了一个清晰的定位。

(1) 系统建立目的:永久保存民族文化遗产、少数民族信息资源的高效管理和应用、提供少数民族事务的发展预测以及未来规划决策帮助等<sup>[6]</sup>。

(2) 系统接受内容的形式以及内容提交的审核工作。常见的内容有:科研论文、预印本、研究手稿、研究报告、技术报告 (文档)、会议论文或会议录、学位论文、专著、项目书、演示资料、机构网站、数字化期刊等多种知识产出或信息资源类型。在提交过程中判断提交文献的完整性、格式校验、是否有病毒感染等,确保收入资源信息系统中的条目质量。

(3) 内容描述及加工过程控制,支持批量导入数据的机制和方式,方便数据的迁移和交换;支持基于 workflow 的数据加工过程管理;支持基于 OAIS 能够选择性搜索、发现、获取和导入外部数字知识库系统中的内容。

(4) 用户对象,主要包括系统管理员、学术团体、研究人员、决策者、普通用户等。另外,对用户定义不同的权限。不同的用户在资源信息系统、社群、专题、条目、元数据和对应的文件上应该有不同的权限。

(5) 服务机构,机构人员 (包括管理员、实验中心、网络信息中心) 对信息的长期保存、网络安全、服务器本身的安全方面的保护,为系统的建立提供必要的设备条件支撑和技术保障。

(6) 服务类型:资源开放存取、资源审核与电子出版发布、资源长期保存、检索查询、信息交流、个性化服务等。

#### 3.2 功能确立

本系统是对数字信息进行管理,将信息保存、服务管理以及人员分配有机地结合在一起,其中包括数字信息采集、信息再生、信息的组织、信息传播以及信息使用和长期保存六个过程。依据 OAIS 参照模型,文中定义系统的框架模型由信息采集系统、信息鉴定及描述系统、工作流系统、信息存储系统以及用户存取系统 5 个部分组成。

(1) 信息采集系统的主要任务是电子文件式数字信息的收集、对传统信息资源数字化同时进行标准化的格式转换等准备工作,然后对整理后的信息进行元数据编辑、存储方式的确定和制定保管计划等系列规范工作。这部分是少数民族信息资源管理的整个系统的主体部分。它能确保信息量以及信息相关元数据的准确性、有效性和完整性。

(2) 信息鉴定及描述系统首先从节省成本角度鉴定哪些信息源需要进行数字化,以及数字化的优先顺序,做到确保信息源质量的同时兼顾完整性和真实性。其次对数字资源形成者提交的信息进行描述,并把文件以标准化的格式著录到数字信息资源系统中。

(3) 工作流系统的主要任务是确保整个系统的数字信息从收集、储存到发布,可以有效正确地运行的整个流程。同时,负责控制系统用户授权和存取控制。对受保护的档案信息进行访问控制,防范用户的越权操作确保系统的安全性;规范用户的提问形式和检索语言,提高用户的检索效率。

(4) 信息存储系统是为数字化信息资源的长期保存和利用做准备的。利用数据字典保存元数据元素的选择和定制;数据交换和共享既要支持 DC 元数据以满足系统提供基于 AI-PMH 协议的元数据开放获取支持,又要支持以 METS 元数据标准作为整个系统中信息包的封装和编码的元数据格式。同时,支持存储海量的数据量、支持存储大数据对象、支持静态和流媒体信息。

(5) 用户存取系统是直接面向用户,具有交互功能。主要任务有:提供可视化服务界面;获取用户检索后所需的文档;自动收集、累计用户的检索需求及分析检索需求。实现分成用户请求、生成响应数据、用户需求分析等几个功能块。

## 4 基于 DSpace 少数民族信息资源系统实现

### 4.1 系统框架

根据 OAIS 的定义,信息包分为提交信息包(Submission Information Package, SIP)、存档信息包(Archival Information Package, AIP)和分发信息包(Dissemination Information Package, DIP)。其中 AIP 是数字信息存储系统的核心逻辑结构。这三个信息包分别对应系统总体设计模型的采集与提交模块、存档与管理模块、发布与服务模块的功能和服务。而各个模块内部又包含若干个子功能模块,这些子模块彼此独立,保持松散耦合,便于系统维护和修改。文中依据需求分析考虑系统知识导航和概念检索需要,提出该系统的功能结构模型。它包括信息存储模块、内容管理模块、系统服务模块三大部分。

信息存储模块是该系统最基本最核心的功能,包括数字资源的管理和存储、资源加工、异构资源的整合、资源调度等关键技术。该模块又分关系数据库和数据流存储管理两部分,关系数据库存储所有资料,包括组织内容、元数据、电子用户和授权、正在运行的工作流的环境,以及维护内容索引等。数据流存储管理部分包括数据流存储、备份和配置数据流存储、配置传统存储、SRB 配置存储等功能。

内容管理模块负责资料管理、使用者管理、权限控制、认证等工作。核心组件用来帮助用户掌握 DSpace 的使用,包括配置管理、常量、背景、电子邮件、日志管理等。内容管理应用程序接口(AIP)可以允许使用的其他类别、修改内存中的存储、元数据、对其他数据格式的支持、包装插件等。插件管理包括插件概念和插件管理的使用。另外,还有工作流系统、管理工具包、电子用户管理、权限管理、句柄管理/句柄插件、检索子系统、浏览子系统、历史记录等组件。

在系统服务模块中,提供了基于 Web 的用户界面应用和管理 DSpace 系统和一系列的接口,便于用户的使用。具体的说,包括网络用户界面、OAI 数据提供模块、条目输入与输出、在 DSpace 实例中传递条目、注册、METS 工具组、媒体过滤器、子社区管理等。

### 4.2 数据模型定义

DSpace 依照其使用机构的需要定义一套“Community-Collection-Item”的数据模型对数据进行组织和管理。本系统 Dspace 系统中成立数个社区(对应于中国地区分布、各省份、56 个民族)和子社区(对应于民族政策、民族历史、民族旅游美食、民族医药、民族音乐、民族语言文字、民族地域及生态保护),每个社区和子社区内可拥有多个数字馆藏(Collection,对应于研究领域、研究项目或技术报告集相关数字文档、数据集、声音视频),而每一个馆藏可能包含数个项目(Item,对应

于具体的数字存档对象,是 DSpace 中实际地封装和管理数字对象的基本单元)以描述其属性,这些项目由 DC 元数据来描述,项目再分为数据束(Bundle),数据束由数字流(Bit Stream)组成,数字流是不可以再划分的、最小的描述单位。每个数字流与一种数据格式相关联。另外,还定义了电子用户(对应于管理员、学术团体、研究人员、决策者、普通用户)和工作流(对应于加入系统的少数民族数字文档进行审核和元数据描述,进而发布信息)。

### 4.3 系统架构政策和原则

政策和原则可以解决系统建设过程中遇到的各种性质不清、责任不明的问题;解决系统在建设和建成后的各种权利与义务关系问题。本系统主要考虑以下几个方面的政策原则:

#### (1) 元数据集中管理策略。

为确保信息源的真实性,对分散数字信息的元数据进行集中管理。首先,制定元数据的管理标准,描述其相关的结构和语义。然后,通过建立元数据交换标准,与外部系统实现无缝集成和互操作。

#### (2) 存储方式的改进策略。

在信息的存储方面,将数字信息分成永久保存数据、长期保存数据和短期保存数据三类。根据信息类型的不同其存储方式也不尽相同。永久保存的数据保存在资源管理存储器中,而短期或长期保管的数据由外部系统负责存储。这样数字信息实体是分别在不同的存储系统中的,并且可以通过元数据进行管理。

#### (3) 权限控制与管理策略。

权限分级及其控制与管理就是区别对待。对象用户不同,权限自然也不同。由于数据库中所包含的内容类型繁杂、格式众多,对不同类型的资源的处理原则会不同。对不同的用户设置不同功能的权限级别,对有特殊要求的资源设定具体的例如时间、范围等的条件,对应不同的分级权限。

## 5 结束语

虽然,DSpace 在开发时已经具有了强大的功能,又提供了许多插件,也支持对已有功能的扩展<sup>[9]</sup>。例如,句柄系统、元数据编辑和开放的 OAI-PMH 接口的功能扩展来支持用户的直接或调整使用;OAI 收割服务扩展实现限定检索、批量获取数字资源、对获取的元数据进行规范化和归并处理等功能。但本项目有其自身特点和不同的需求,这些特殊的功能需要通过定制和补丁来解决,这是下一步的工作重点<sup>[10-12]</sup>。

少数民族决策支持系统是建立一个庞大的复杂工程,没有现成的固定模式,所以技术要求高,需要投入

(下转第 234 页)

亮度范围。此外,经过测试,用户通过界面进行的亮度范围设置能够正确发送给控制节点,实现不同阈值下的差异照明控制。

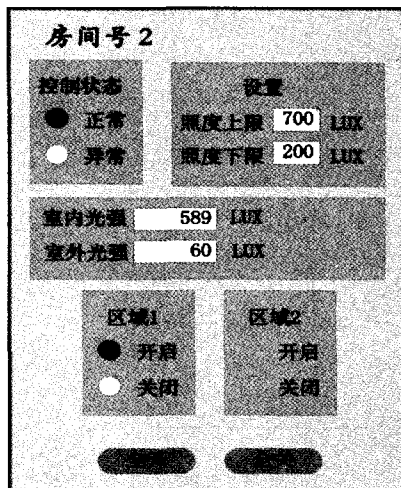


图 6 照明控制界面

## 5 结束语

文中设计实现了一种综合 IPv6 无线传感器网络和红外控制技术的新型室内照明控制系统,该系统能够依据用户设置的亮度阈值、环境光强及现场人员位置,智能裁决并控制照明设备开关,并且支持用户远程监控系统工作状态。

文中详细介绍了系统的硬件结构和软件设计流程,着重分析了终端控制节点采用特定编码进行设备控制器的实现原理,并详细说明了功能验证的效果。该系统使照明控制智能化,有利于节约能源、延长照明设备使用寿命,并且管理方便、维护开销有限。本系统

具有很强的扩展性,可以大规模部网实现楼宇智能照明控制,也可以增加传感器类型进行其他设备的自动控制,因此该系统具有良好的发展前景,为物联网技术应用与智能建筑行业作出有益探索。

## 参考文献:

- [1] 张公忠. 物联网与智能建筑[J]. 智能建筑与城市信息, 2011(1):14-17.
- [2] 孙利民,李建中,陈渝,等. 无线传感器网络[M]. 北京:清华大学出版社,2005:5-10.
- [3] 储昭勋,胡艳军. 无线传感器网络技术[J]. 计算机技术与发展,2006,16(4):64-66.
- [4] 霍宏伟,张宏科,部帅,等. 一种 IPv6 无线传感器网络节点的设计与实现[J]. 计算机应用,2006,26(2):303-306.
- [5] 陈庆章,赵小敏,毛科技,等. 以无线传感器网络构建的照明节电系统的设计[J]. 机电工程,2010,27(3):1-3.
- [6] 彭东. 一种传感模块无线传感器网络新型节点的设计与实现[D]. 北京:北京交通大学,2009.
- [7] 田洪强,秦雅娟,郑涛,等. 无线传感器网络智能红外控制节点的实现[J]. 计算机应用,2010,30(9):2249-2252.
- [8] TEXAS INSTRUMENTS. CC2420 Datasheet[EB/OL]. 2007-05. <http://focus.ti.com/cn/lit/ds/symlink/cc2420.pdf>.
- [9] Lian Xiaoqin, Hu Jianbin, Zhang Xiaoli, et al. Design and Implementation of Indoor Environmental Monitoring Wireless Sensor[C]//2010 3rd International Conference on Advanced Computer Theory and Engineering. [s.l.]:[s.n.], 2010.
- [10] 叶琳. 基于 Web 的远程监控信息管理系统的设计与应用[D]. 上海:上海交通大学,2009.
- [11] 周晓伟,蔡建平,郑增威,等. 新型室内照明智能控制系统的研究与实现[J]. 计算机应用研究,2009,26(8):2977-2981.

(上接第 229 页)

大量的信息资本。本课题只是其中的一个环节,而要想实现最终的决策支持,还必须继续在建设知识组织体系、研究型用户个性化工作空间的建设、策略的规范、版权研究以及对用户的使用情况进行统计分析等方面进行深入的研究。

## 参考文献:

- [1] 杜昌钰,龙飞. 基于 DSpace 少数民族资源信息系统构建[J]. 信息与电脑,2010(3):170-174.
- [2] 吴江华. 开放性档案信息系统:背景、职责及功能[J]. 图书情报知识,2006(5):85-87.
- [3] 王军. 基于成本分析的数字资源长期保存策略研究—迁移法与仿真法比较[J]. 图书情报知识,2006(1):74-77.
- [4] DSpace System Documentation[EB/OL]. [2007-07-05]. [http://www.dspace.org/index.php?option=com\\_content](http://www.dspace.org/index.php?option=com_content).
- [5] DSpace System Documentation: Functional Overview—Data mode[EB/OL]. [2007-07-05]. [http://www.dspace.org/index.php?option=com\\_content&task=view&id=149#data\\_model](http://www.dspace.org/index.php?option=com_content&task=view&id=149#data_model).
- [6] 蔡华利. 数字资产管理系统—DSpace 研究[J]. 情报科学,2006,24(9):1405-1408.
- [7] 孙常丽,付佳,石丹. DSpace 系统的应用现状简述[J]. 科技情报开发与经济,2008(21):79-80.
- [8] 杨武健,王学勤. DSpace 机构知识库系统的分析与研究[J]. 现代情报,2006(11):220-222.
- [9] 谢静,王军. DSpace 数字仓储简介及其应用分析[J]. 数字图书馆论坛,2007(9):18-25.
- [10] 王伟军. 基于 DSpace 的企业知识服务平台建设初探[J]. 信息系统,2008,31(9):920-923.
- [11] 张蓓,董丽,李新伟,等. 基于 Fedora 的数字资源管理方案的研究与实现[J]. 现代图书情报技术,2005,123(5):1-5.
- [12] 祝忠明. 基于 DSpace 构建学科知识库系统的研究与实践[J]. 数字图书馆,2006,139(7):10-14.