

# 基于 DSpace 系统的特色数据库建设实践<sup>\*</sup>

金松根 孙常丽 王琳琳

(牡丹江医学院图书馆 牡丹江 157011)

〔摘要〕 在深入介绍 DSpace 系统的框架、安装过程和应用软件基础上, 结合牡丹江医学院图书馆实践, 阐述影像学特色数据库平台的搭建, 包括存储文献类型、具体建库步骤, 并对系统定制和数据库管理进行探讨。

〔关键词〕 特色数据库; DSpace; 标准化

**Practices of Constructing Characteristic Database Based on DSpace System** JIN Song - gen, SUN Chang - li, WANG Lin - lin, Library of Mudanjiang Medical College, Mudanjiang 157011, China

〔Abstract〕 Based on in - depth description of the framework, installing process and application software of DSpace system, with the combination experience of Library of Mudanjiang Medical College, the paper elaborates the construction of imaging characteristic database, including the types of document storage, concrete steps of constructing the database, discusses the system customization and database management.

〔Keywords〕 Characteristic database; DSpace; Standardization

目前大多数高校图书馆除自建硕博学位论文数据库、重点学科导航库这些传统模式的数据库外, 都自主开发建设了结合本校学科特色、地域特色的数据库。特色数据库建设的目的是把形式多样、来源广泛或特殊收藏的、关于某个主题的数字化资源进行组织整合, 形成有序的、便于利用的特色资源集合。以“建设特色资源, 展示学术成果, 长期知识保存”为宗旨, 根据学院学科发展实际情况, 2009 年牡丹江医学院图书馆以普通高等学校第一类特色专业建设点, 即影像学院学术成果为特色数据库内容, 启动了“影像学特色数据库”建设项目,

2009 年底基本建设完成 DSpace 系统架构和部分类型文献的数据输入, 并发布到校园网上投入试用。

## 1 DSpace 系统平台

### 1.1 系统概述

DSpace 系统是 MIT 和 HP 实验室合作研究开发的, 为教育科研机构创建一个稳定的环境来长期保存和获得原生数字材料——捕捉、存储、索引、保存以及传播数字资产的系统平台。DSpace 系统管理和发布由数字文件或“位流”组成的数字条目, 并且允许创建、索引和搜索相关的元数据以便定位和存取该条目, 满足长期稳定的保存和管理的数据库建设需求。

### 1.2 系统安装

选用类 Unix 操作系统 (Redflag Linux6.0) 来

〔收稿日期〕 2010 - 04 - 09

〔作者简介〕 金松根, 馆员, 主任, 发表论文 10 余篇。

〔基金项目〕 黑龙江省教育厅人文社会科学项目“基于 DSpace 影像学特色数据库建设研究”(项目编号: 11542256)。

搭建 DSpace 系统。使用类 Unix 操作系统是由于其服务稳定、编写的代码移植性强、对开源软件提供良好的支持和文件系统符合 DSpace 需要等原因。由于 DSpace 源代码是由 HTML 和部分嵌入的 Java 代码组成的,还必须下载 Java 开发包 JDK1.5 和 Apache ANT 1.7.0 以保证 DSpace 的编译和运行环境。Web 服务器则采用 Tomcat 6.0。后台的数据库使用的是默认的 PostgreSQL 8.2 (其更高版本存在兼容问题)。

DSpace 的安装步骤简单介绍如下:从 <http://www.dspace.org> 下载 DSpace 的源代码;下载并配置 JDK, ANT, Tomcat 和 PostgreSQL;根据 Tomcat 和 PostgreSQL 的设置配置 DSpace 的配置文件 `dspace.cfg`;利用 ANT 编译 DSpace;使用命令 `create-administrator` 创建管理员账户;发布 DSpace 到 Tomcat。具体安装步骤见《基于 DSpace 的学位论文管理系统的开发》一文<sup>[1]</sup>。

### 1.3 系统编码标准化

基于 Unicode/UTF-8 的统一编码环境,设定 DSpace 基于 Java 技术设计和开发,其国际化主要通过 Java 技术在支持国际化软件设计和开发方面的特性和优势,采用 Unicode 标准和可本地化独立资源文件的技术和方法来实现。

在典型的基于 Java 的 Web 应用环境中,彻底地解决不同编码格式转换这一问题的主要方法是统一各环节的编码格式,建立基于 Unicode 标准的统一编码环境。DSpace 即采用了这一方法来实现对多语言环境的支持和本地化。以 DSpace 使用 Tomcat 6.0 同时作为 Web 服务器和应用服务器为例,要保证以 Unicode/UTF-8 方式对 Http 请求数据进行编解码,只需在其配置文件 `server.xml` 中增加如下有关 URIEncoding 的设置<sup>[2]</sup>:

```
<!-- Define a non-SSL HTTP/1.1 Connector on port 8080 -->
```

```
<Connector port = "8080"
```

```
protocol = "HTTP/1.1" connection Timeout = "20000"
```

```
redirect Port = - "8443"
```

```
URIEncoding = "UTF-8" />
```

对于任何 DSpace 中的 JSP 文件,都在其第 1 行

使用 JSP Page 指示符明确指定字符编码格式为“UTF-8”。在安装数据库环节,使用 PostgreSQL 8.2 版本,在创建供 DSpace 使用的数据库时也要使用编码选项并声明为“Unicode”,如“`createdb -U dspace -E Unicode dspace`”,这样就使数据库内部以 UTF-8 方式对数据进行编码和存储。

## 2 数据库建设

### 2.1 存储文献类型

DSpace 以支持程度把上传到系统的文献类型分为:Supported, Known, Unsupported 3 级。其中完全支持的文献即 Supported 类型文献有:marc, pdf, ps, mpa, abs, mpeg, wav, jpeg, jpg, txt, xml 等文本、图像、声音、视频类型,完全满足特色数据库所需的科技期刊、专利、技术报告、学位论文、会议录以及电子图书、课件、光盘镜像等文献的上传。

### 2.2 建库步骤

以徐敏等人发表的论文“肾动脉狭窄的超声、MSCTA、DSA 影像检查对比分析”为例,详细介绍 DSpace 系统下的数据录入步骤<sup>[3]</sup>。

在 <http://192.168.30.222:8080/jspui> 网址登录,输入提前申请的 E-mail 地址和密码。进入“MY DSpace”,点击“Start a New Submission”。选择专题,点击下拉菜单可以看到专题列表,点击“已发表论文”选项。在提交页面的顶部,会看到代表提交进程的 7 个椭圆形按钮链。随着提交过程的继续,相应的按钮颜色也发生改变。在这些已经改变颜色的按钮中,可以随意点击任何一个按钮进入相应的提交进程,在这过程中,所提交的数据不会丢失。

开始提交论文条目时,会出现以下几个选项:

(1) 多个题名——有时论文条目有一个以上的题名,可能是题名缩写、首字母缩写、翻译题名等。若论文条目有这种情况,则选择此项。(2) 曾经发行过——以前未曾出版或发布的新论文条目提交到 DSpace 时,系统将分配一个发布日期。已经发表的

文章选此项。(3) 多个文件——在 DSpace 中论文条目可以由多个文件组成, 如一篇文章附录有视频和数据文件。若要提交这样的论文条目, 选择此项。点击“Next”按钮继续或点击“Cancel/Save”按钮来中止或取消提交。

在接下来的两个页面中, 填入的信息将成为论文条目的元数据记录。部分描述信息, 见表 1。

表 1 部分主数据描述信息

| 项目   | 具体描述                           | 项目      | 具体描述                             |
|------|--------------------------------|---------|----------------------------------|
| 作者姓名 | 徐敏                             | 引用      | 中国医疗设备;<br>2009/11               |
| 其它作者 | 刘春风; 李明珠                       | 系列编号    | 无                                |
| 题名   | 肾动脉狭窄的超声、MSCTA、DSA<br>影像检查对比分析 | 类型      | 论文                               |
| 其它题名 | 如翻译题名或缩写题名等, 则可填入              | 语言      | 中文                               |
| 发行日期 | 2009-11-15                     | 主题词或关键词 | 肾动脉狭窄; 彩色多普勒超声诊断; 影像诊断           |
| 出版者  | 中国医疗设备出版社                      | 文摘      | 剪切/粘贴方式或手工打字的方式输入论文条目的文摘, 文摘字数不限 |

上传文件, 有两种方法填入要上传的文件名称: 一是直接在输入框中填入文件的完整路径和文件名。二是点击“浏览”按钮, 在弹出的窗口中选择相应的文件并双击, 相应的文件名即被填入到输入框中。当上传完文件后, 需要检查是否上传正确。点击文件名, 可以在新的窗口中打开文件, 这时可以检查内容是否正确。检验核实论文条目的描述信息, 同意机构仓储发布协议许可文件, 完成提交该论文条目到数据库中, 提交完成。

为了避免毫无价值的材料混入特色数据库中, 影响特色库质量以及用户的利用, 必须建立质量控制机制。对全过程进行质量控制, 包括建库前、建库中、建库后。其办法主要是组织有关人员和数据进行审核、核对。如果不组织专人进行认真、严格的审核, 数据库的质量是很难保证的。因此, 建库后的审核也是质量控制不可缺少的一环<sup>[4]</sup>。

3 结论

从特色数据库建设的实际情况来看, 数据库的

框架虽然已经构建, 但面临的问题还有很多, 有待于进一步解决。首先要加强认识和理解 DSpace 系统使用的开放源代码软件及软件包。在实际应用中, 应用层程序调用业务逻辑层和存储层程序来实现用户所需要的各种功能, 因此对存储层主要的软件包 org. dspace. storage. rdbms. \*、org. dspace. storage. bitstore. \*, 应用层主要的软件包, 即 org. dspace. app 目录下的所有软件包要有深入了解和认识<sup>[5]</sup>。熟悉 DSpace 系统中应用的开源软件 Lucene, PDFBOX, Tm-extractor, OAICat 等会对修改定制 DSpace 系统大有好处。其次要认识到只重开发、不重使用是违反专题文献数据库建设效益论的本质的。开发人员不但要建设高质量的数据库, 还要对数据库的使用做跟踪调查。调查该数据库是否能满足用户的需求、检索是否简单易用、还存在哪些问题和不足、改进的方向是什么、宣传力度是否足够、是否涉及知识产权问题等, 跟踪调查能保证数据库不断改进、完善。

总之, 特色数据库建设是网络时代赋予图书馆的重要任务。高校图书馆应以专业化服务为己任, 深入研究信息技术, 进行有效管理, 争取学校政策支持, 稳步、高质量地推进特色资源数据库建设发展。

参考文献

1 金松根, 孙常丽. 基于 DSpace 的学位论文管理系统的开发 [J]. 中华医学图书情报杂志, 2008, 17 (4): 59-61.

2 [http://www.dspace.org/1\\_5\\_2Documentation/DSpace-Manual.pdf](http://www.dspace.org/1_5_2Documentation/DSpace-Manual.pdf) [EB/OL]. [2010-01-10].

3 祝忠明, 马建霞, 常宁, 等. 基于 DSpace 构建学科知识库系统的研究与实践 [J]. 现代图书情报技术, 2006, 139 (7): 10-14.

4 田丽丽. 专题数据库建设的质量控制 [J]. 图书馆学研究, 2007, (7): 23-25.

5 <http://wiki.dspace.org/confluence/display/DSPACE/DSpaceResources> [EB/OL]. [2010-01-10].