

赵洗尘

论数字图书馆专题资源库的建设

摘要 数字图书馆资源建设的主要方式是共建共享。在合作建设、资源共享的模式下,服务对象的确、专题的选择、总体设计以及相关标准与规范的确是建设数字图书馆专题资源库的关键。图3。参考文献5。

关键词 数字图书馆 专题资源 资源建设
分类号 G250.76

ABSTRACT Cooperative development and resources sharing is the major pattern of the development of digital library. In this paper, the author thinks that key factors in the development of subject databases include: determination of customers, selection of subjects, general designing and drafting of related standards and specifications. 3 figs. 5 refs.

KEY WORDS Digital library. Subject resources. Resource development.

CLASS NUMBER G250.76

William Y. Arms 在其《数字图书馆概论》第一章中指出:“数字图书馆可非正式地定义为有组织的信息馆藏及相关服务,信息以数字化形式保存,并通过网络进行访问。定义的核心在于说明信息是有组织的。”也就是说,如何组织信息资源,是数字图书馆建设的核心。本文结合“中国少年儿童信息大世界——网上图书馆”建设,对此进行探讨。

1 专题资源库建设的典型过程

数字图书馆中资源库的建设比通常所说的数据库的建设要复杂得多,不但要考虑如何存贮各种类型的数字化资源,还必须充分考虑资源与资源之间的相互关系、资源之间的共享问题以及如何在资源发布过程中实现相关资源的链接与导航。图1是一个典型的专题资源加工流程。

在整个过程中,选择与确定专题是关键,总体设计是整个过程的核心,是专题资源库质量与服务的保障。

2 选择与确定专题

2.1 服务对象的确定

确定用户对象范围,即确定服务的读者群。对少年儿童图书馆而言,读者对象不但包括0~8岁的孩子,还有成人,而且不同年龄段孩子的阅读需求与习惯差别很大,所以明确服务对象,就成为专题资源建设时首先要考虑的因素。

2.2 专题资源调查

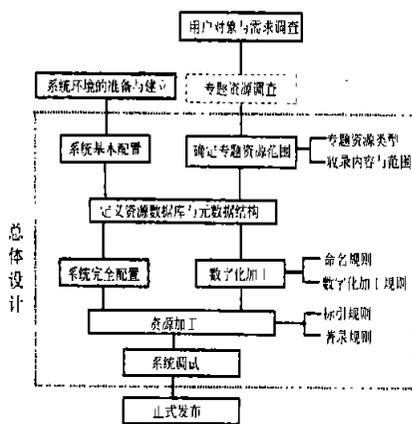


图1 专题资源加工的典型过程

对合作馆藏文献资源、网上资源与社会资源进行综合性调查研究,并充分考虑在资源搜集、选择过程中存在的障碍,以保证资源的系统性、完整性与实用性。

2.3 专题确定原则

充分考虑用户的实际需要与阅读习惯、阅读能力、阅读方式是确定专题的基本原则;强调专题内容的思想性、科学性、知识性与趣味性;注重专题内容的优选性、特色性、层次性;保证专题内容的长期性、稳定性与发展性;保证专题内容的系统性、完整性与实用性。从以上原则出发,深圳少年儿童图书馆开发的“中国少年儿童信息大世界——网上图书馆”(以下简称少儿网上图书馆)确定了“我爱五星红旗”、

“少儿科技园”、“少年文苑”、“闪光的足迹——名人故事”、“育儿专栏”、“家教园地”、“童话世界”、“再塑一个你——少儿礼仪”等多个专题。

3 选择与确定资源范围

选择与确定专题资源范围是指确定专题资源的形式与内容。在综合考虑少儿信息资源特点的基础上,少儿网上图书馆确立了选择与确定资源范围总则:(1)多媒体原则:收录多种媒体的资源,如文本、图像、声像、视像资料;(2)多素材原则:同一内容不同形式,如全本、少儿本、节本、画册、图片、网上内容综合收录;(3)精选原则:不同版本、不同出处,但内容相同、形式相同,精选一个较完整的版本或内容,或根据具体情况进行综合编辑;(4)多渠道原则:资源来自馆藏资源、网上资源、社会资源、自编辑资源等;(5)综合原则:综合编辑同一内容不同媒体的资源,如综合不同出处的文字与图片资源。

对于每一个具体专题,则根据专题资源的具体特点,确定具体的资源选择细则。

4 总体设计

总体设计包括:专题资源库结构、链接关系、标引规则、数字化规则以及加工流程设计。

4.1 资源库整体结构与元数据结构

数字图书馆中的专题资源库建设,不仅仅是将专题资源数字化,更注重对多种媒体与多种形式专题资源在共享原则下的重组与整合。因而,专题资源库从总体结构来说可分为三层,见图2。其中数字化资源层是基础资源;数据库层是主体部分;专题资源库层实际上是个虚拟层。

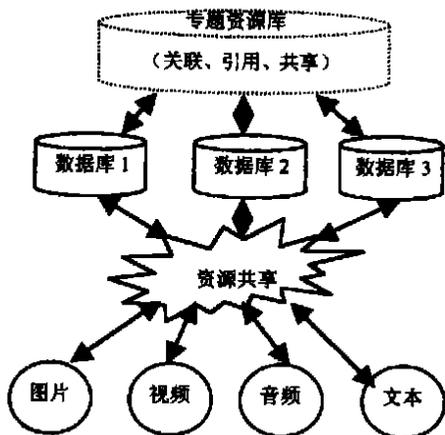


图2 专题资源库的整体结构

专题资源库由多个数据库组成。专题资源由多种媒体与形式的资源组成,数据库就是根据这些资源的特点以及资源的表现形式而确定,如“再塑一个你——少儿礼仪”专题资源库由“礼仪全文”、“礼仪故事”、“专题书目”这三个数据库组成。

数据库的元数据结构,其设计也是总体设计的一个重要组成部分。一般情况下,将资源描述字段按树型结构组织,并尽可能采用相关的国家标准或国际标准,如都柏林核心集等。

4.2 数据库之间的关系

数据库之间是相互关联的,它可以通过三种方式实现:一是通过检索途径将所有数据库联系起来,二是在数据库之间建立字段级的联结,三是通过置标,即在全文字段中为相关词语加标识,通过标识词链接到相关记录。

图3简单示意了“再塑一个你——少儿礼仪”资源库中三个数据库的主题与索取号之间的链接关系。

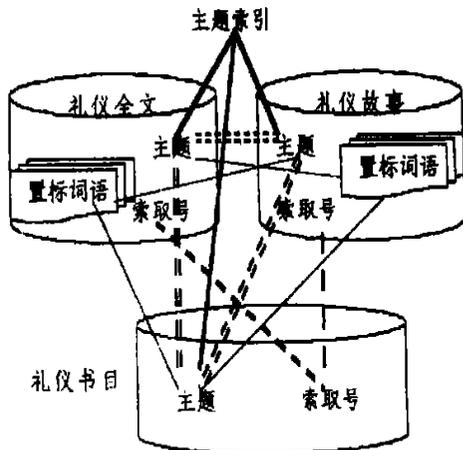


图3 资源数据库之间的关系

主题索引串联起三个数据库中的所有主题字段;三个数据库中主题字段的联结使每个主题词在发布界面中可直接点击检索;礼仪全文与礼仪故事中的索取号与礼仪书目中的索取号相联,使用户直接点击可以查看书目信息;礼仪全文和礼仪故事全文或提要字段中置标后的词语直接作为主题词进行检索。于是,相对独立的三个数据库通过索引、字段联结和置标紧密联系在一起,将三个树型结构组织的资源变成了一个网状的、相互关联的整体。

4.3 标引规则

标引的目的不仅在于对资源内容的揭示,更要从用户使用、用户兴趣等多个角度来考虑。尤其对

少年儿童用户,激发其兴趣,追求简单易用,力求使用户的使用过程既是知识组织系统又是循序渐进的导读过程,既考虑主题标引的规范性又兼顾主题词或关键词的通俗性,这些都是制定标引规则时必须考虑的重要原则。标引规则主要包括:著录规则、主题标引规则、分类标引规则与置标规则。

著录规则指所有数据库中各数据字段的著录内容与内容选取规则。传统的著录规则适用的对象及著录要求与专题资源著录有很大差异,因而不能完全适应专题资源库建设的需要,必须在其基础上制定具有针对性又简洁适用的著录规则。

主题标引规则定义主题标引用词以及标引深度。对于专题资源的主题标引,我们采用关键词标引,所选用的关键词,概念清晰、简练,科学性、专指性强,对描述和表达资源的主题起关键作用。

针对专题资源的特点,我们没有采用现有的分类法,而是从分类的角度为读者提供一个专题资源的主题分类表,使用户能够顺其体系了解专题资源的全貌。此系统的分类,不使用类号而直接用类目名称;不局限于一种体系,可按多种分类标准同时建立多种体系,如按学科分类同时再按地域分类等。在设计专题资源的知识组织系统时必须考虑:不同专题资源的知识组织系统不同;类目设计应兼顾导读功能;等级层次不宜过多,且应保持平衡;类目名称的设定应兼顾既能吸引人,概念又明确;类目词概念清晰,具有检索意义。

置标是将全文中的词加标识,以方便用户使用。在数字资源加工系统中,被置标的词是全文中出现的在某个索引字段中著录的词,如主题字段中著录的关键词、分类字段中著录的类目词、人名字段中著录的人名等。在制定置标规则时首先要确定所置标的类型,如主题标识、分类标识、人名标识等,然后确定为哪些词置标。置标不能盲目,如果很多置标词下没有链接内容,会引起诸多不良影响。

4.4 数字化规则

数字化规则是指对各种介质资源进行数字化加工时必须遵守的规则。根据少儿资源的特点,数字化规则主要包括:文件命名规则、图像处理规则与全文文本处理规则。

在“少儿礼仪”资源库的建设中,我们初步确立了以下几个命名规则:(1)文件名的结构为:“索取号_所在页码_区分号”或“ISBN_所在页码_区分号”或“ISSN_年卷标识_所在页码_区分号”。(2)由于“数字资源加工系统”对介质文件名最多识别8位,则装入系统时将文件名按知识组织系统中的类目名称缩

写代码命名,即改为:“一级类名代码_二级类名代码_区分号”。(3)同一内容的不同介质的文件名用相应的代码标识,如:“t-图片”等。

少儿文献中图片、图画特别多,而且形式多样。如连环画有影视剧照、黑白画、彩色画等。年久画册,纸黄,两面透字,扫描质量差。为保证图像清晰和网上传输效率,建立一套有效实用的图像处理规则是必须的,这些规则包括:规定扫描过程中的各项参数的设置;规定图像处理流程,如锐化次数等;规定图像大小;规定图像格式。

为保证数据的安全性,我们同时以文件与全文内容两种方式装入资源库,对扫描识别后的文本内容如下:

文件格式:.txt 文件格式。

文件内容:按原文内容,不加工标识。

5 结束语

数字图书馆的根基在于实实在在、持续不断的资源建设,而不仅仅是模式、架构、系统的研究。国内外数字图书馆资源建设的主要方式是共建共享。在合作建设、共享资源的模式下,专题选择、总体设计以及相关标准与规范的确定是关键。对于数字图书馆资源库的建设,尤其是面向少年儿童用户资源库来说,简单地将资源扫描识别,简单数字化是不够的,按一定的知识体系将资源重组,建立资源间的关系网是至关重要的。

参考文献

- 1 William Y. Arms 著;施伯乐等译. 数字图书馆概论. 北京:电子工业出版社,2001
- 2 北京现代文津信息技术研究中心. 数字资源加工系统
- 3 中国少年儿童信息大世界—网上图书馆管理委员会. “中国少年儿童信息大世界—网上图书馆”专题资源库建设总体方案,2001
- 4 Bernard J. Hurley et al. The Making of America Testbed Project: A Digital Library Service Model. The Digital Library Federation. Council on Library and Information Resources, December 1999
- 5 Gail Hodge. Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files. The Digital Library Federation. Council on Library and Information Resources, April 2000

赵洗尘 深圳少年儿童图书馆副研究馆员。通讯地址:深圳市红荔路。邮编 518027。

(来稿时间:2001-09-07)