

彝族口述历史资料数据库的资源体系构建研究

高建辉^{1,2}

(1.楚雄师范学院图书馆,云南 楚雄 675000;2.云南大学,昆明 650000)

摘要:彝族口述历史资料具有重要的历史文化价值,但这些珍贵的资源及其载体正在快速流失和消亡,急需进行数字化保护。该文首先阐述了构建彝族口述历史资料数据库资源体系的意义,然后研究了彝族口述历史资料数据库建设过程中资源体系构建的原则和方法,最后得出了彝族口述历史资料数据库的资源目录体系和元数据信息表。彝族地区的图书馆、档案馆、文化馆等机构在进行彝族口述历史资料采集和数据库建设时,可以依据这一框架和目标,开展彝族口述历史资料的收集和数据库建设工作,实现资源的共建共享。

关键词:口述历史资料;数据库;资源体系

中图分类号:G250.74;K281.7 **文献标志码:**A **文章编号:**1673-1883(2019)02-0001-04

Study on Creation of a Resource System for Databases of Oral Accounts of Ethnic Minority Yi's History

GAO Jianhui^{1,2}

(1. Library, Chuxiong Normal University, Chuxiong, Yunnan 675000, China;

2. Yunnan University, Kunming, Yunnan 650000, China)

Abstract: The oral accounts of ethnic minority Yi's history have important historical and cultural values, but these precious accounts and their carriers are rapidly dying and vanishing, and urgently need to be preserved in digital forms. This paper first expounds the significance of creating a resource system for databases of oral accounts of ethnic Yi minority's history, then studies the principles and methods for the creation, and finally develops the resource catalogue system and metadata information table of the oral history database. The libraries, archives, cultural centers and other institutions in ethnic Yi regions can launch the collection and database construction work for oral accounts of Yi people's history in terms of this framework and goal, so as to share efforts and resources.

Keywords: oral account of history; database; resource system

彝族是中国第六大少数民族,有着悠久的历史 and 丰富多彩的民族文化,很多极具特色的彝族非物质文化遗产都是通过口耳相传的形式流传至今。随着信息技术的发展以及国家对少数民族历史文化保护工作越来越重视,部分彝族地区的图书馆、文化馆和档案馆等机构都开始了彝族口述历史资料的收集工作。但就目前来看,这些机构搜集到的彝族口述历史资料都呈现出类别各异和种类单一的特点,比如云南省民委只收集了彝族的民间故事、史诗歌谣和神话传说,且都只有文字记录,彝族地区的很多文化馆主要收集了彝族非遗传承人的口述资料 and 教学片,这样的资源结构不利于彝族口述历史资料的系统性保护,且难以实现资源的共建共享。其原因主要是学术界对这一问题研究较少,

对于搜集的内容还没有一个统一的标准。为尽快建立彝族的口述历史资料数据库,并实现不同机构资源的共建共享,需要科学合理地确定其资源结构和资源体系,从而指导相关机构开展资源的搜集、整理和建库等工作。

一、构建彝族口述历史资料数据库资源体系的意义

第一,资源体系的构建可以使得彝族口述历史资料数据库的资源结构更加科学、合理,便于资源的整合与优化配置。首先,彝族口述历史资料数据库中的资源结构是否科学、数据层级设置是否合理、内容的涵盖范围是否全面是评价其质量的重要因素,因此,在数据库建设前需要经过充分的调研,

收稿日期:2019-03-15

基金项目:国家社会科学基金项目:西南彝族口述历史资料搜集整理及其有声数据库建设(16XTQ008)。

作者简介:高建辉(1982—),男,云南楚雄人,副研究馆员,博士研究生,研究方向:图书馆学、历史文献学、数字化与数据库建设等领域。

尽量全面搜集不同类别的数据,综合相关搜集机构的意见来合理构建资源体系。其次,如果有了统一的数据库资源分类方法和结构体系,资源搜集机构就有了统一的搜集目标,在这一框架下来协作开展资源建设工作,在不同机构间实现资源的合作共建,避免重复劳动,浪费人力和资金。

第二,资源体系的构建是开展数据库资源建设工作的理论基础和前提。数据库的建设包括平台建设和资源建设两部分,后者是建设的核心,包括资源的采集、价值鉴定、分类和著录等工作。由于彝族口述历史资料的资源内容具有多样性和不确定性,这些具体的工作都要以资源体系作为基础。资源体系对资源的采集范围和资源结构的框架进行了限定,对于没有民族学、图书馆学和计算机科学背景的建设者也可以依据该体系进行资源的搜集和整理工作。从而为资源采集工作指明方向,简化资源建设的流程,提高资源建设的效率,保证资源的质量。

第三,资源体系的构建有利于彝族口述历史资料数据库建成后的开发利用。首先,资源体系可以作为数据库门户网站的资源导航,按类别或类型直观展示数据库的内容,在数据库宣传时,可以使用户方便快捷地了解数据库的主要内容。其次,每条数据搜集的主要信息都是检索点,资源体系的确定可以确保主要检索点信息的完整性。最后,资源体系的构建可以使数据库的资源的层次清晰、逻辑严密,所有数据成为一个有机的整体。在资源开发利用时,便于构建知识关联和语义关联,挖掘数据深层次的价值,提高知识服务的质量。

二、构建彝族口述历史资料数据库资源体系的方法

(一)资源体系构建的原则

一是资源体系结构的科学性。资源体系的结构要存在内部逻辑,形成有机整体,并具有可扩展性。同一层次的类目应当相互排斥,突出各自的性质特征,内涵清楚明确,外延界限分明,便于资源的归类。每一层次的分类中对同一上位类必须使用同一个分类标准,逐层确定分类标准,使其兼顾科学性和适用性。资源体系层次少了就不能反映资源的外延范围,多了就会使管理变得复杂化,一般设置三层左右比较合适,体现三种主要的分类方法,其他分类方法可以在数据库的元数据信息中体现^[1]。

二是资源体系的系统性和完整性。资源体系必须形成一个完备的整体,不论过去、现在和将来

采集到的所有资源都可以找到归属类目。资源的内容要涵盖口述历史本体资源和相关资源,本体资源即口述者的口述录音、录像或转录文字稿,其他资源包括与口述者和口述内容相关的图片、图书、文章、口述者个人情况、项目记录文档等文献资料^[2]。数据条目是构成数据库的基本单元,每一条数据信息的完整性是数据库资源体系完整性的基础,因此,每条数据所包含的元数据类型也需要健全完整,涵盖尽可能多的类别信息。

三是资源内容的独特性和价值性。首先,资源体系的内容要立足于彝族历史文化的保护和传承,具备彝族特色,便于彝族文化的传播和展示。其次,资源体系的内容要符合口述历史的特点,重视搜集民间记忆,记录普通百姓的生活,弥补文献档案的缺失和不足。另外,本体资源的内容要符合口述历史资料的内涵,必须包含使用人类语言口述的内容,不能只有纯器乐音频或纯舞蹈视频^[3]。最后,在进行体系中各类资源和层级目录的设置时,要考虑该类资源的数量和相对价值,不能把所有类别都罗列上去,数量少且内容接近的类目可以合并,数量极少的类目可以省略并使其包含在其他类目的下一级类目中。

(二)资源体系构建的步骤

根据彝族口述历史资料数据库资源体系构建的原则和目标,在具体设计时,需要经过以下四个步骤。

第一步,查阅文献,实地调研,收集资源。通过查阅相关资料,对彝族口述历史资料的概念、内涵和外延进行界定,明确口述历史资料的来源、种类、特点和价值,根据来源收集各类型的彝族口述历史资源,研究资源的形式和内容特点,掌握资源的总体情况。

第二步,研究资源的分类方法和资源类型。对搜集到的资料进行分类统计和分析,研究其分类方法,统计不同分类方法下各个类目的资源数量。笔者对楚雄师范学院图书馆收集到的三千多条彝族口述历史资料进行了研究,结合数据库平台建设的要求,得出彝族口述历史资料常见的单一分类方法有十一种,分别是来源分类法、表现形式分类法、地点分类法、口述对象属性分类法、记录形式或载体类型分类法、口述语言分类法、方言或支系分类法、时间分类法、口述者姓名分类法、通用学科主题分类法、彝族特色主题分类法。计算机存储的本体资源类型包括文字、音频、视频,其他资源类型包括图书、报刊、论文、图片、音视频等^[4]。

第三步,确定数据库的层级和结构,合理设置类目和属类。根据各类资源的规模和数量以及数据库平台建设的要求,数据库的资源体系分为三层,逻辑结构采用资源分类体系,体系中主要的分类方法是主题分类法、口述者分类法和数据类型分类法。既可以采用以口述人物为主的分类方法,先按口述者姓名分,再按主题和类别分,使得所有口述者的相关口述质量相对集中,也可以采用以专题为主的分类方法来构建资源分类体系,突出资源的内容属性和价值,保持资源内容之间的联系。

第四步,多方征求意见,在实践中不断修正资源体系,使其类目趋于完整和健全。在数据库建设的前期工作和资源搜集过程中,可以根据所搜集到的资源动态调整资源体系的结构。同时,在设计资源体系的过程中,还需要听取民族学、计算机科学、语言学 and 图书情报学专家的建议,征求相关收集机构的意见,不断改进方案。根据彝族口述历史资料的特点,主题分类法是比较适用的,但这一方法的类目设置与资源归类难以掌握,必须经过反复研究、实践、补充、修订分类方案,随着资源的不断增加使其趋于稳定。

三、彝族口述历史资料数据库资源体系的构建方案

(一)数据库的目录体系设计方案

资源体系设计的科学性、合理性与适用性将直接影响资源的归类,从而决定数据库检索和利用的效果。根据以上原则与方法,利用“通用学科主题—彝族特色主题—数据格式”这一资源分类结构设计彝族口述历史资料数据库的资源目录体系(表1)。

(1)一级目录。首先,在第一级目录中采用了通用学科主题分类法,并做了模块化处理,一共分成七个模块。如果某一主题的下一级目次太少,则归入或合并到相近主题,比如历史和政治合并,语言与文学合并等,这样的合并使得主题类别更加简洁明了,便于资源归类。其次,一级目录中没有经济、军事等主题。其原因是彝族口述历史资料中很少有专门针对这两个方面来表现的内容,这符合口述历史的特点,而且这两方面在文献历史资料中记载的比较多,两者形成互补,如果有机构搜集到的这两类资源数量较多,也可以增加相关一级目录。第三,“相关资源”是这一分类体系中比较综合与特殊的一类,其包含了本体资源以外的相关内容,是彝族口述历史资料不可缺少的组成部分。

(2)二级目录。在第二级目录中采用了民族特

色主题分类法,省略了部分资源数量较少的彝族特色资源类型,比如彝族克智数量较少,其数据并入语言文字。这样处理会使得彝族特色会降低,但通用性变强,更容易归类。其次,歌谣原本是属于艺术的下位类,既二级目录,但由于彝族的歌谣数量众多,在彝族群众的生活中起着非常重要的作用,彝族的很多历史文化都是以歌谣的形式传唱至今,因此,把歌谣提出来与艺术并列作为一级目录,以突显其重要性。

(3)三级目录。在第三级目录中,采用数据类型分类法,本体资源的每一个二级目录都对应着文字、音频和视频三种类型,同一条资源的不同数据类型可视为不同资源进行处理。“相关资源”则对应着文字、音频、视频、图片、图书、报刊、论文、网页等众多数据类型。第三级目录之所以采用这一分类法,原因是为了让数据库前台页面针对不同数据类型

表1 彝族口述历史资料数据库的资源体系

一级目录	二级目录	三级目录
历史与政治	州县乡村史	文字
	民族发展史	
	个人及家族史	
语言与文学	社会评论	音频
	神话传说	
	创世和叙事史诗	
歌谣与艺术	语言文字	音频
	民间故事	
	表演艺术	
	酒歌	
科技与教育	生活及劳动歌	视频
	礼仪习俗歌	
	情歌	
	生产生活技艺	
哲学与宗教	医药卫生	视频
	天文历法	
	学习与教育	
礼仪与习俗	理想信念	视频
	道德伦理	
	宗教祭祀	
相关资源	人生礼仪	文字、音频、视频、图片、图书、报刊、论文、网页等
	节庆习俗	
	生产生活习俗	
	禁忌与习惯	
	讲座会议	
	学术资料	
	影视剧集	
	宣传介绍音视频	
	相关图片	

的资源按其特点设计不同的页面进行显示。比如对于视频资源,需要在三级页面设计视频播放器,如果是图书,则可以在二级页面增加图书封面图片。

(二)数据库的元数据设计方案

经过多年的研究和积累,借鉴其他数据库的元数据设计方案,得出了彝族口述历史资料数据库的元数据设计方案,如表2所示。每一个字段名都标明了其设置的原因和作用,大多数的信息都反映了彝族口述历史资料的主要分类方法和重要检索点。该表只是列出了本体资源的元数据,对于“相关资源”中的图书、报刊、论文等资源已有比较成熟的元数据体系,直接借鉴即可。任何机构如果想收集彝族口述历史资料并建立数据库,都可以直接参照这一元数据设计。此外,在开展具体的标引著录工作时,还要制作一个元数据著录标准和要求,以便统一和规范每个字段的著录格式,使最终的著录

信息能够批量导入数据库。

四、结语

彝族口述历史资料数据库的资源体系是在楚雄师范学院图书馆建设“西南彝族口述历史有声数据库”的实践基础上总结出来的,所提出的理论和方法具有较强的实用价值。如果彝族地区所有的保护机构在建设彝族口述历史资料数据库时都采用这一资源体系,就可以为数据库的共建共享建立基础。同时,希望相关问题的研究能够引起政府和文化机构对彝族口述历史资料保护工作的重视,积极开展彝族口述历史资料的收集和数据库建设工作。这一体系对于其他少数民族口述资料数据库的建设也具有较强的借鉴价值,构建出各民族的口述历史资料数据库资源体系,最终建成中国少数民族口述历史资料数据库。

表2 彝族口述历史资料数据库本体资源的主要元数据信息表

字段名称	设置原因
题名	区别于其他数据的一句话,是该条数据最重要的检索点
内容提要	逐字记录和翻译的反映原始资源主要内容的一段文本,用于快速了解该条数据的内容
关键词	选取主要内容中涉及的1~3个实词,作为主要检索点
日期	记录资源采集的时间
责任者	对数据内容负责的人,明确资源的著作权
版权所有	拥有版权的个人或机构,明确资源的版权
来源	保持资源与来源之间的关系,记录资源的流转情况
采访者	永久记录采访人的姓名,便于按采访者检索
口述者	永久记录口述人的姓名,便于按口述者检索
演唱(奏)者	永久记录演唱(奏)人的姓名,便于按演唱者检索
表演者	永久记录表演人姓名,便于按表演者检索
资源格式	计算机中保存的资源文件的扩展名,便于按格式归类和检索
栏目	该条数据所属二级目录,便于按目录名称检索
语种	彝语、汉语、英语、彝汉
方言类型	东、西、南、北、中部等方言、次方言、土语
主题分类	该条数据内容所属的主题,可以体现省略或合并了的主题信息
地点	资源收集或录制地点
时长(分钟)	音视频录制的时长
比特率(Kbps)	音视频录制的质量
分辨率	视频录制的尺寸
文件大小(MB)	在计算机中保存的该资源占用的存储空间
整理上传者	数据整理和上传到数据中的人,一般是技术人员,在数据上传前要进行数据的规范性检查和处理
全文链接	音视频文件的下载链接地址
封面	用作该条数据在网站前台显示的图标
口述者详情	口述者的基本情况。包括民族、年龄、籍贯、职称职务、是否毕摩或非遗传承人、传承项目等
备注	其他需要标注的字段

参考文献:

- [1] 华林. 西南少数民族历史档案管理学[M]. 北京:民族出版社,2001:166-167.
- [2] 吴雨遥,成灵慧.民族记忆抢救的彝族侗族人口述档案资源建设研究[J].兰台世界,2017(23):12-15.
- [3] 赵局建,周燕.石林彝族口述档案的收集原则[J].兰台世界,2015(20):11-12.
- [4] 楚雄师范学院图书馆. 西南彝族口述历史资料数据库[EB/OL]. [2019-02-11].http://yzksls.cxtc.edu.cn:8081.