

机构规范文档构建方式研究*

王星¹, 曾建勋¹, 苏静², 贤信¹

(1. 中国科学技术信息研究所, 北京 100038; 2. 武汉大学信息管理学院, 武汉 430072)

摘要: 探索机构规范文档的构建方式, 可有效解决信息检索、计量评价等活动中机构名称著录混乱、层级结构模糊的瓶颈问题, 为我国机构规范文档建设提供借鉴。通过调研机构规范文档的应用需求以及理论、实践研究现状, 在梳理机构名称及其属性特征、机构实体关联关系的基础上, 构建机构规范文档元数据框架, 设计机构规范文档构建流程。结果表明, 机构规范文档的构建流程包括机构数据获取及规范化处理、文档记录的审核以及面向不同需求的应用设置。同时, 应从机构规范文档的场景需求分析、智能构建方式、协同共建机制和共享应用能力四个方面共同推进机构规范文档建设工作。

关键词: 机构数据; 规范文档; 元数据框架; 协同共建

中图分类号: G254

DOI: 10.3772/j.issn.1673-2286.2015.07.001

机构泛指高等院校、医疗机构、公司企业、科研院所和政府机关等实体对象, 该类对象数量庞大、层级关系复杂, 且名称变化形式多异、衍化多样。新兴机构的出现, 传统机构的淘汰、更名、重组与合并, 名称拼写形式的不规范, 缩/简写名称、惯用名称与首选名称交错使用等诸多现象导致相关或相似机构名称大量存在, 给编目著录、信息检索等活动造成严重困扰, 继而影响检索效率以及统计分析、计量评价的可信度。因此, 有必要引入规范文档 (Authority File) 对机构名称进行关联化、系统化研究。其中, 规范文档的概念在文献编目领域由来已久, 是指借助编目参照方法集中所选用标目与未选用名称的记录文档, 其作用是实行规范控制来确保标目的唯一性和稳定性^[1-2]。薛明则进一步指出, 规范文档是把正式采用的标目形式与个人或者团体责任者、题名、主题等其他未被采用的名称形式, 通过参照方法集中起来所做的全面记录和说明^[3]。陈金星等基于规范文档提出名称规范方法, 即通过建立字典式的数据表对不同形式责任者的名称标识进行穷举式列举^[4]。由此, 大数据环境下的机构规范文档可理解

为涉及机构名称的规范化描述、识别、聚类、应用等多方面的综合性记录文档, 其构建方式是指在数字信息规范控制过程中, 通过收集、识别、匹配与归一机构记录, 将同一机构所有名称形式以及其他属性信息有序集中, 并赋予选用机构标目唯一标识号, 以达到梳理机构层级关系、组织结构及其属性特征之目的。

1 国内外机构规范文档的建设实践与理论研究

1.1 国外方面

规范文档的思想最早出现于1876年美国著名图书馆学家卡特编制的《印刷本字典式目录条例》, 1978年《英美编目规则》(第二版)则明确说明建立规范档的详细规则, 引发编目界重视^[5]。而规范文档的建设首创于美国国会图书馆, 也称为规范表, 专指规范记录的有序集合^[6]。2001年欧洲十余家文化机构参与开启“国际规范资料库合作建制计划”(Linking and Exploring

* 本研究得到国家自然科学基金面上项目“基于海量数字资源的科研关系网络构建研究”(编号: 71273251)资助。

Authority Files, 简称LEAF)^[7], 其目的为通过收集名称规范文档的分布式搜索系统, 创建同一个体或团体不同名称形式相互关联的数据结构, 以实现基于互联网的网络链接与检索共享^[8]。2003年美国国会图书馆、德国国家图书馆和OCLC共同提出的构建虚拟国际规范文档项目(Virtual International Authority File, 简称VIAF)则是解决机构名称规范问题的有益尝试: VIAF通过统一处理各机构提供的原始数据生成规范记录, 并为之分配唯一标识符, 且提供维护规范文档的合作机构链接地址, 大大降低图书馆界的维护成本^[9]。截至2014年7月, VIAF创建的规范数据已包含5百多万条团体记录^[10]。

在标准化制定方面, 2008年美国国家信息标准组织(National Information Standards Organization, 简称NISO)成立机构标识符工作组(NISO I2), 负责创建一套泛在数字信息需求下稳定、可扩展并能实现互操作的机构标识标准。与此同时, 国际标准化组织研究并制订国际标准名称标识符(International Standard Name Identifier, 简称ISNI), 即ISO27729:2012。经过多方评估与协商, 最终于2013年推荐采用ISNI作为机构唯一标识标准。2012年, Ringgold成为ISNI的首个注册机构, 并将其应用于Ringgold标识数据库, 经过全面归一处理机构数据, 该数据库已构建完成机构的详尽数据列表与清晰的层次结构^[11]。

值得注意的是, 现阶段的机构规范文档之间已实现一定程度的关联应用, 如VIAF引入ISNI记录, ORCID通过Ringgold标识数据库获得ISNI机构标识符、Ringgold机构标识符、机构规范名等, 汤森路透与Ringgold的机构规范数据进行无缝链接等^[12]。

此外, 不少学者就大数据环境下机构规范文档构建的理论研究层面提出许多可行方案, 主要涉及机构名称的规范化描述、识别、聚类、应用环境等内容。例如, 日本庆应义塾大学学者Kimura在2013年通过比较中日韩三国对中国机构名称的描述差异, 创建了跨语言形式的多重检索^[13]; Jiang等人采用基于归一化压缩距离(Normalized Compression Distance)的机构名称聚类方法, 实现同一机构多种名称的高效集成^[14]; 美国宾夕法尼亚州立大学的Yoojin Hong等人通过建立人物规范文档赋予每个作者唯一ID号, 当作者所在的机构名称发生变动时, 将当前名称作为规范名, 旧名作为属性字段予以保存, 并在相应作者名称标识之间建立参照, 实现人物区分和机构延续功能^[15]。

1.2 国内方面

21世纪初, 我国信息服务机构已逐步认识到规范工作的重要性。国家图书馆、CALIS联机合作编目中心、香港地区以及北京地区的高校图书馆都纷纷加入机构规范文档建设中, 如名称规范库构建、规范规则探索、规范系统研究等^[16]。此外, 国家图书馆、台湾汉学研究中心、CALIS与香港地区大学图书馆协作咨询委员会联合创建中文名称规范联合数据库检索系统, 涉及个人名称、团体名称、会议名称、丛编题名的规范记录。中国医学科学院医学信息研究所通过计算机辅助和人工审核来实现医学领域机构著录的形式规范、消歧与关系规范, 其建立的机构规范库包括规范机构名81676个、别名314192个, 涉及医疗机构、高等院校、国家/省部级/中科院重点实验室等。中国科学技术信息研究所着力加强基于引文的知识链接数据库建设, 其中的机构名称库涵盖机构规范名称与非规范名称对照、机构基本属性、机构上下级隶属关系、机构中英文名称对照以及机构名称变迁等五类数据^[17]。万方数据公司则完成收录20多万家企业机构、科研机构、信息机构和教育机构的基本信息, 包括机构类型、曾用名、负责人、成立时间、简介、通信地址等, 借此整合学术论文、科技成果、专利、标准等机构科研产出。

同时, 各种机构名录数据库对机构信息进行不同程度的描述, 如全国组织机构代码信息数据库是依托《全国组织机构代码编制规则》国家标准, 向我国境内依法注册、登记的机构依次分配九位组织机构代码号, 形成包含地址、电话、邮编、经济行业、经营范围等信息的机构规范文档; 中国国家认证认可监督管理委员会除描述机构名称、电话、地址、网址等内容外, 还对其分支机构进行关联; 伊梅资源有限公司整理的机构名录数据包括负责人、电话、网址等机构联系方式, 以及产品服务、行业类型、市场规模等信息。

此外, 国内机构规范文档的理论研究也取得阶段性成果。在机构名称的数据描述方面, 张建聪等通过复用DC元数据集以及自定义核心元素等对机构要素加以定义, 并利用RDF/XML构建面向知识导航的机构元数据互操作模型^[18]; 在机构名称识别与聚类方面, 沈嘉懿等人在对中文组织机构全/简称特征进行分析的基础上, 提出一种基于规则识别中文组织机构全称和简称的方法, 并应用于中文关系的抽取系统^[19]; 胡万亭等人提出了一种基于词频统计的组织机构名识别方法^[20], 孙海

霞等借鉴现有基于频繁词集的文本聚类初始中心确定方法,基于K-means提出一种面向大规模机构名称规范处理应用的机构聚类方法,并构建机构规范名-别名映射表辅助构建工具^[21];在机构名称应用环境方面,曾建勋等以学术论文为例,探讨面向知识评价的规范文档建设,对比分析知识评价和资源组织中的规范文档构建差异,提出知识评价中的机构规范文档需要保留机构的标准名称及其层级关系^[22]。

综上所述,虽然我国在中文机构名称规范化建设方面已有一定基础,但相较于国际机构规范文档开放共享工作的有序开展,我国还处于机构规范数据的基本属性描述和多源集成的起步阶段,尚未实现机构名称的自动监测、关联更新与规范管理,缺少机构规范文档的系统性研究,亟待开展机构规范文档的可持续建设与研究工作。因此,本研究以学术论文发文作者所在机构为研究对象,探索机构规范文档的构建方法以及推进策略,为解决机构名称著录混乱、关联层级模糊等问题寻求自动化、流程化、规范化的解决路径。

2 机构规范文档构建的方法研究

通过研究大数据环境下科研机构实体的命名规则、演化规律、属性特征及实体间的关联关系,设计机构规范文档元数据框架,并将之应用于学术论文数据库建设,提出合理的机构规范文档构建流程。

2.1 机构名称及其属性

(1) 机构命名规则

与普通名词相比,机构名称隶属于专有名称,主要具有两个特点:一个是名称结构较为灵活,长度不固定;二是名称结构比较紧凑,不带从句和“的”等助词。从语言学角度来看,它是一种偏正式复合名词,由一个或者一个以上的定语加上表示机构称呼的中心语组成。其中,定语可分为前后两部分,前一部分主要包括国名、地域名、方位词、人名、序数或基数词、专造词等表示地理位置、级别的外部特征信息,后一部分与机构的活动内容相关,表达机构的所属行业、涉及学科、服务产品和创办方式等,中心语则表明机构的性质或类别信息,如学校、企业、医疗机构等^[23]。

(2) 机构名称演化规律

机构名称的演化与机构变革的关联程度较大,一

是机构自身体制、层次的根本性变化或者利益驱使导致的机构更名,二是机构合并、兼并、拆分重组等涉及多个机构的名称变异。以近年来高校名称的演化规律为例,一是中心语从“学校”到“学院”再到“大学”的升级转型,二是地域特征由“小”变“大”,如由“省”到“华中”再到“中国”,三是过时、过度专指的词汇向流行、综合性词汇过渡,如去除“农业”、“纺织”、“冶金”等专业词汇,增添“科技”、“工程”等当下的流行词汇。

(3) 机构属性特征

不同类型的机构所具有的属性存在差异性,针对不同需求任务对机构属性研究的侧重点也会有所不同,一般可以从机构类型、地域、行业、学科、水平、层级、产出等方面对机构属性进行研究。其中,机构类型可依据机构功能或性质,将其分为行政机关、事业单位、公司企业、社会团体以及其他组织机构五大类^[24]。针对机构地域、行业、学科等属性进行分析时,可借助国家地区代码、经济行业代码以及学科分类表等信息组织工具,保障数据规范性与灵活性。此外,水平属性是指机构按同一标准进行资质考核评定的等级或水平。层级属性是指某一机构在所涉及的纵向机构体系中的位置或者层次说明,如一级机构、二级机构、部属机构、省属机构等。

2.2 机构实体关联关系

在社会网络活动中,任一机构实体都不是独立个体,必然与其他机构实体存在错综复杂的关联关系。通过对机构实体的多种属性揭示、标识和规范,可形成语境化的机构实体多维网络,有助于提高检索系统中应用机构名称的查全率与查准率,提升统计分析与评价的可信度,有助于系统梳理机构实体之间的层级关系,辅助机构分级及快速锁定目标机构。机构实体关联关系具体分为机构内部关联和机构外部关联,前者是指某一机构实体自身产生的关联关系,包括单一机构实体各个名称之间的关联关系和机构整体与其内在各部门的上下级关系,如用代关联、参照关联、属分关联等;后者是指多个机构实体通过某种共性或者活动而产生的关联关系,如地域关联、行业关联、学科关联、合作关联、产出关联、隶属关联、附属关联等。需要说明的是,属分关联、隶属关联和附属关联三者之间的区别界定是以该机构是否具有独立法人资格为基准:若

机构整体具有独立法人资格, 与其不具备独立法人资格的下属各部门之间的关系定义为属分关联, 与其具有独立法人资格的下属机构之间的关系定义为附属关联, 与其具有独立法人资格的上级管理机构之间的关系定义为隶属关联。

2.3 机构规范文档元数据框架

为满足不同情境下用户的使用需求, 需遵循一体性、互操作性、可扩展性等原则, 在复用都柏林核心元数据元素及其扩展版DCMI元素的基础上, 设计基于唯一标识符的机构规范文档元数据框架。机构规范文档描述性元数据元素主要包括机构标识符、规范名称、交替名称、曾用名、机构类型、地域、简介、主题等。其中, 机构标识符 (dc:identifier) 是指系统自动分配的机构唯一标识字符串, 具有永久标识和持久链接的特性; 规范名称 (dc:title) 是指机构实体多个名称中的首选名称, 可直接应用全国组织机构管理中心的注册登记名称; 交替名称 (dcterms:alternative) 是与规范名称共同指代同一机构实体的其他名称形式, 包括译名、别名、简称、俗称等; 主题元素 (dc:subject) 可从行业 (Industry)、学科 (Subject catalog) 等方面进行扩展自定义, 并依据国家标准《中华人民共和国学科分类与代码国家标准》、《国民经济行业分类》限定元素值域。

除描述性元数据外, 还需定义管理型元数据, 如声明机构名称经过规范化处理的日期、方式、责任人等, 并可从创建日期 (dcterms:created)、变更日期 (dcterms:modified) 两方面扩展日期 (dc:date) 元素。

此外, 在描述机构实体关联关系 (dc:relation) 时, 规范名称与交替名称之间的关系称为用代关联 (use)、与曾用名之间的关系称为参照关联 (dcterms:references), 属分关联包括包含 (dcterms:hasPart) 与被包含 (dcterms:isPartof) 两种关系, 机构实体与其前身机构、后继机构可通过替代关联 (dcterms:replaces)、被替代 (dcterms:isReplacedby) 关联表示。机构规范文档元数据的具体元素及其关联关系如图1所示。

2.4 机构规范文档构建流程设计

以海量学术论文数据库为研究对象, 基于机构规范文档元数据框架, 设计一套富有实践性、扩展性、智能性的机构规范文档构建流程。目标一是半自动识别、匹配不同来源的机构名称及其属性数据, 并阶段性更新维护其变更状态、对其关联关系进行揭示与管理控制; 目标二是面向大众提供机构注册与机构信息校正服务, 形成机构规范文档的开放式利用与可持续发展机制。

具体而言, 机构规范文档的构建流程主要包括机

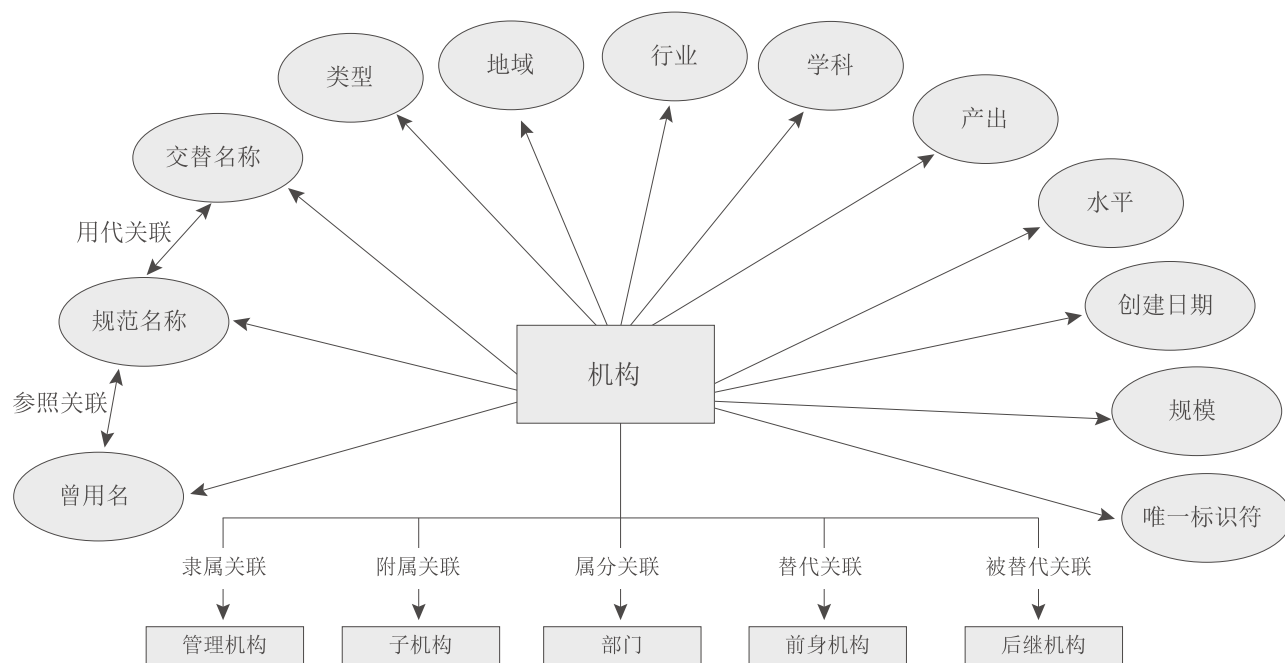


图1 机构规范文档元数据框架示例

构数据的获取、处理、审核与应用四个环节,如图2所示。首先,机构数据的获取与处理,是以学术论文数据库中的“作者单位”数据为基础,依次切分、提取、分类机构数据,并对机构名称进行识别、消歧、匹配以及机构属性的标注,规范处理后能够有效揭示机构的关联关系,从而构建机构树及机构规范文档。同时,通过集成全国组织机构代码信息数据库、机构名录数据库等其他机构数据来源,利用用户注册与登记方式,可提升机构规范文档的内容丰富度和用户参与度。此外,联合应用国际较为通用的机构唯一标识符系统,为机构规范文档中的机构规范名称赋予数字资源标识码,更便于与其它系统的互操作。其次,机构规范文档的数据审

核是保障相关服务产品推广应用的前提,应成立机构规范记录审核团队负责评估机构数据来源、校正人工审核记录以及处理用户注册记录等。最后,面向不同用户群体的特定需求,形成专业程度高、安全性能强、可迁移的机构规范数据集应用服务产品,以提高用户规模 and 用户粘性。

3 机构规范文档构建的推进策略

3.1 加强机构规范文档的场景需求分析

海量信息环境下,机构规范文档构建是以多元异构

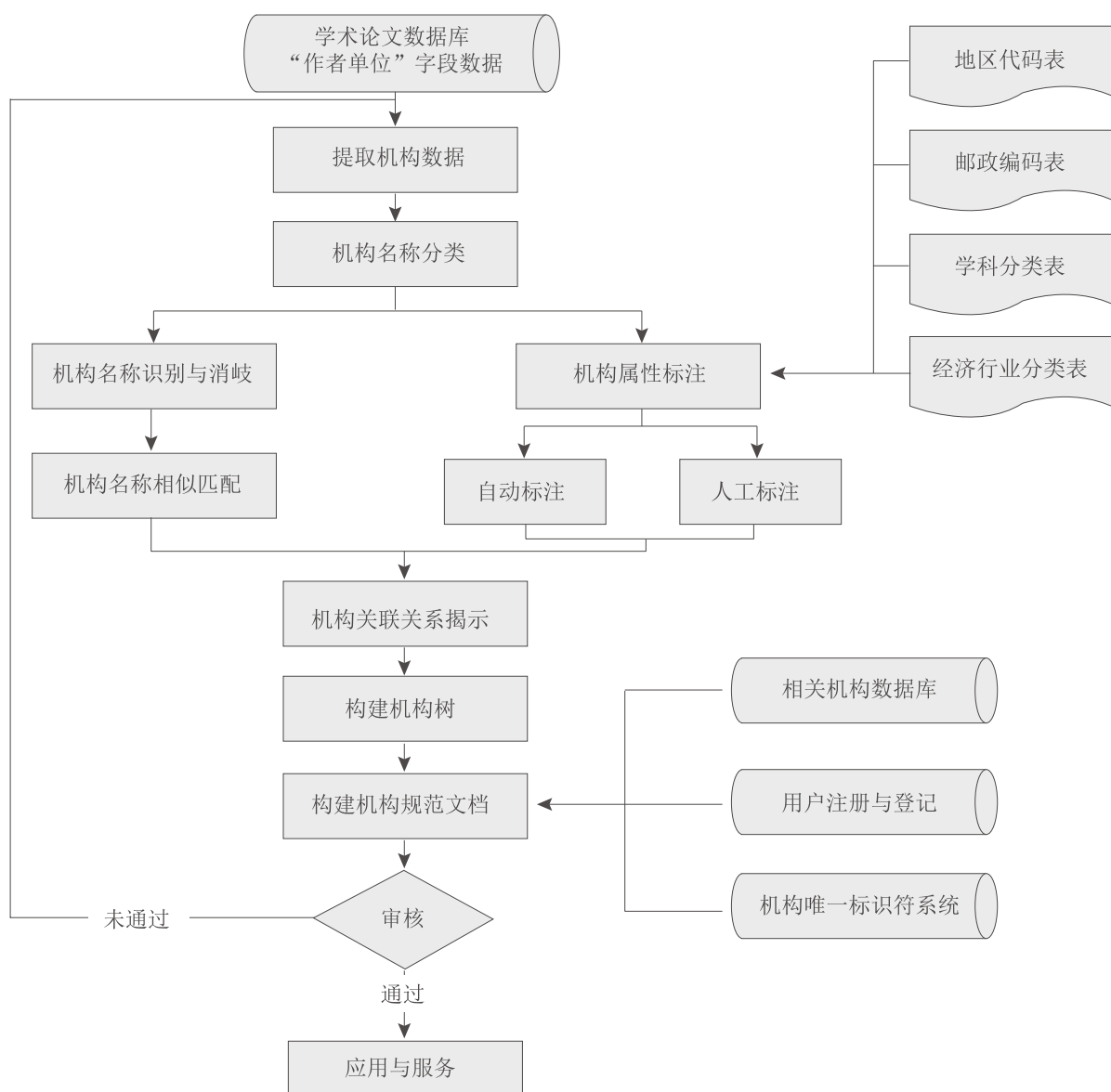


图2 机构规范文档构建流程

的机构数据仓储为基础,应根据用户使用需求定制目标数据并优化服务内容,以保障机构规范文档的高效应用。因此,机构名称规范控制需细化机构规范文档的使用场景,如信息检索、计量评价、团队发现、趋势分析等,并针对不同使用场景的功能与目标,分析其对机构名称规范控制的需求情况,包括数据来源选择、元数据元素定义、运行实施方案制定、产品服务形态设计等。例如,计量评价与团队发现对机构规范文档元数据的需求便存在差异性,前者更为强调机构名称描述的全面性、准确性及其结果的归一化,力求高效辨别相似性较强的机构实体名称、保障集成同一机构实体的所有有效记录;后者相对侧重机构实体属性的多维度描述,用户可基于自身状况和实际需求的评估结果,根据机构所在地域、所属学科、行业背景、科研产出等重点特征,寻找匹配度较高的目标合作团队。

3.2 完善机构规范文档的智能构建方式

目前,人工审核是机构规范文档构建的必要路径,需与自动算法相结合完成机构实体名称识别与消歧、机构属性标注、社会公众登记等关键环节,以符合机构规范文档的质量标准。但是,爆炸式增长的大数据环境要求机构规范文档建设应以提供智能解决方案为目标,以保证机构规范文档数据质量为前提条件,通过改进机构规范文档的自动化构建技术,尽可能控制人工成本。例如,不断优化机构数据识别、消歧、匹配等规则与算法,以提升高相似度的机构名称并发检索和实时数据分析的性能要求;定向监测权威网站的机构数据,识别并调用、集成新增的机构名称及关联信息,以补充或更新规范文档资源;添加机构规范数据自动抽查程序,定期生成机构数据录入、处理等记录报告,以检测一段时间内数据异常现象。

3.3 强化机构规范文档的协同共建机制

机构规范文档协同共建机制是指在统一制度约束下,采取分布式机构信息资源构建方式,充分发挥各方主体优势和协作能力,以保证信息资源内容的多样性、完整性和一致性。首先,在追踪国际机构名称规范化研究、实践工作以及相关标准制定的基础上,联合图书、情报、档案、出版等领域专家,提出适用于我国机构规范文档建设的元数据标准,以此推动规范文档的有效

产生、管理和应用。其次,明晰共建主体的职责范围和参照规则,统一共建主体的机构名称处理能力,包括制定机构规范数据集的流程化数据集成方法、标准化数据维护与更新原则等。再次,大力提升包括社会公众、研究人员、相关机构在内的参与意识和鼓励措施,使其积极融入到机构规范文档的构建流程,如机构原始数据的提供、机构规范文档数据的校正与完善等。

3.4 提升机构规范文档的共享应用能力

机构规范文档的构建目标之一,即提升机构规范文档的资源共享与服务应用能力,拓展服务对象和服务范围,为不同用户群体提供高质量机构规范数据集的信息组织服务。这就要求,一是提升面向用户的机构规范文档系统服务功能,根据用户需求提供分层次的服务内容,如机构实体信息简单查询服务、机构实体信息延伸性认证服务、特定主题的机构信息专项咨询服务等,二是开发支持系统嵌入的机构规范文档对外接口,如面向社会公众的机构数据注册与校正、面向图书馆、出版商等信息服务主体的关联数据迁移与交换、面向创新主体的机构团体发现与认证等。此外,随着我国机构规范文档构建体系的稳定发展,应努力与国际通用、认可的机构唯一编码等数字对象标识系统相融合,以扩大机构规范文档的共享应用范围。

4 结语

机构规范文档是解决机构名称著录混乱、层级关系结构模糊的可行方法,并已在国际范围内得到有效验证。我国应积极借鉴国际先进成果开展机构规范文档的研究和实践,以填补现代信息资源建设过程中的不足,提升机构信息服务质量。机构规范文档构建是一个长期持续更新、维护与优化的过程,需要机构实体、研究人员、社会公众等各方在内的积极参与和共同努力,才能保障机构规范文档的统一建设、稳定运行和实践发展。

参考文献

- [1] 沈宝顺. 中文文献数据库中的规范档建设[J]. 上海高校图书情报学刊, 1995, 5(1): 34-36.
- [2] 黄俊贵. 规范控制概说[J]. 高校图书馆工作, 1999, 19(3): 1-8.

- [3] 薛明,王丽萍. 我国规范文档研究综述[J].图书馆学刊,1999,21(1):28-30.
- [4] 陈金星,祝忠明. 责任者名称规范控制研究及进展[J].现代图书情报技术,2009(12): 12-17.
- [5] 薛明,王丽萍. 我国规范文档研究综述[J].图书馆学刊,1999,21(1):28-30.
- [6] 彭斐章. 目录学教程[M].北京:高等教育出版社, 2004:208.
- [7] 贾君枝,石燕青. 中文名称规范文档与虚拟国际规范文档的共享问题研究[J].中国图书馆学报,2014,40(6):83-92.
- [8] 王芳,王小丽. 基于OAI协议的数字档案馆元数据互操作问题研究[J].现代图书情报技术,2007(3):18-24.
- [9] 崔春,毕强. 虚拟国际规范文档(VIAF)项目进展[J].图书情报工作,2014,58(6):129-134.
- [10] OCLC.2014 Annual Report to VIAF Council[R/OL]. (2014-08-13)[2015-05-23]. <http://www.oclc.org/content/dam/oclc/viaf/OCLC-2014-VIAF-Annual-Report-to-VIAF-Council.pdf>.
- [11] Ringgold. Identity Database[EB/OL].[2015-05-27]. <http://www.ringgold.com/identify>.
- [12] 窦天芳,张成昱,张蓓等. ResearcherID现状分析及应用启发[J].图书情报工作,2014, 58(4):40-45.
- [13] Kimura M. Differences in Descriptions of Chinese Personal and Corporate Name Authority Data: A Comparison between China, Japan and South Korea[J]. Library and Information Science, 2013(69):19-46.
- [14] Jiang Y, Zheng H T, Wang X, et al. Affiliation disambiguation for constructing semantic digital libraries[J]. Journal of the American Society for Information Science and Technology, 2011,62(6):1029-1041.
- [15] Hong Y, On B W, Lee D. System support for name authority control problem in digital libraries: Open DBLP approach[M]//Rachel H, Liz L. Research and Advanced Technology for Digital Libraries. Germany: Springer Berlin Heidelberg, 2004:134-144.
- [16] 刘丽静. 中文名称规范数据库的建设与发展[C]//北京高校编目研究会.北京高校编目研究会2007年学术年会论文汇编.2007:14-19.
- [17] 王立学. 国家工程技术图书馆知识服务模式探究[J].数字图书馆论坛,2013(10): 26-31.
- [18] 张建聪,吴广印. 面向知识导航的机构要素元数据规范及互操作[J].情报学报,2010,29(1): 84-92.
- [19] 沈嘉懿,李芳,徐飞玉等. 中文组织机构名称与简称的识别[J].中文信息学报,2007,21(6):17-21.
- [20] 胡万亨,杨燕,尹红凤等. 一种基于词频统计的组织机构名识别方法[J].计算机应用研究,2013,30(7): 2014-2016.
- [21] 孙海霞,李军莲,吴英杰,等. 基于k-means的机构名聚类研究与应用[C]//中国医学科学院医学信息研究所, 图书馆2011年学术年会论文集, 2012:60-65.
- [22] 曾建勋,王立学. 面向知识评价的规范文档建设方法[J].图书情报工作,2012,56(10): 101-106.
- [23] 曾文婷,曾建勋. 科学文献中机构要素词典的构建[J].情报杂志,2010, 29(4):121-125,154.
- [24] 中华人民共和国国家标准.GB/T 20091-2006 组织机构类型[S].北京:中国标准出版社,2006.

作者简介

王星,男,1977年生,中国科学技术信息研究所工程师,研究方向:信息管理与信息系统。

曾建勋,男,1965年生,研究员,博士生导师,中国科学技术信息研究所信息资源中心主任,研究方向:信息资源建设。

苏静,女,1988年生,武汉大学信息管理学院博士生,研究方向:知识组织与语义出版, E-mail: owensujing@163.com。

贤信,女,1990年生,中国科学技术信息研究所硕士生,研究方向:信息资源建设。

Research on the Construction of Institutional Authority File

WANG Xing¹, ZENG JianXun¹, SU Jing², XIAN Xin¹

(1. Institute of Scientific and Technical Information of China, Beijing 100038, China;

2. School of Information Management, Wuhan University, Wuhan 430072, China)

Abstract: Exploring the construction of institutional authority file can effectively solve the bottleneck problems in the activities of information retrieval and measurement evaluation, which organization names are described confusedly and its hierarchy structure is fuzzy. By investigating the application requirements and practical research situation, it analyzes the name and property of institutions and its relationship. Moreover, it designs the metadata framework and construction process of institutional authority file. The results show that the construction process of institutional authority file should include the acquisition and normalization processing of institution data, the audit of document record and application settings for different requirements. At the same time, we should promote the construction work of the institutional authority file from the requirement analysis, the intelligent construction methods, the collaborative construction mechanism and the ability to share applications.

Keywords: Institution Data; Authority File; Metadata Framework; Collaborative Construction

(收稿日期: 2015-06-21; 编辑: 雷雪)