

科技人才信息资源的分类编码设计 ——以《中国高层次科技人才数据库》建设为例*

□ 吴晓莉 王运红 / 中国科学技术信息研究所 北京 100038

摘要: 文章在优化、整合、更新、补充现有的各种与科技人才信息相关的分类编码的基础上, 提出一套基于科技人才信息的属性特征, 同时兼顾人才基本信息、知识和技能、科技活动、科研成果产出及科研诚信等信息的分类编码方案。该方案能有效地组织、关联和揭示科技人才信息资源, 对科技人才信息资源数据库的建设、数据共享具有重要作用。

关键词: 分类, 编码, 科技人才, 人才信息

DOI: 10.3772/j.issn.1673—2286.2012.10.009

1 引言

2006年国家主席胡锦涛在全国科技大会上宣布2020年中国将建成创新型国家, 使科技发展成为经济社会发展的有力支撑。近年来, 我国科技人才数量增长迅速, 2008年全国科技人力资源总量已达496万人, R&D人员总量达到196.5万人, 科学家和工程师总量达到159.2万人^[1]。因此, 面对海量的科技人才信息, 政府、高校、院所、科技企业和中介机构都已开展科技人才信息化、标准化建设工作。

人才信息的标准化的建设、科技人才数据库的建设工作, 是实施人才战略、建设人才队伍以及全面推进高层次人才队伍建设的基础。立足现在, 着眼长远, 利用现代信息技术建成统一管理、结构合理、技术先进、功能完善的科技人才管理信息系统, 全面提升国家各个人才管理部门的人才管理、开发和服务工作水平, 充分开发和利用人才

资源, 为国家的各项事业的发展合理地配置人才资源提供人才信息的保障。科技人才数据库是中国高层次人才管理的基础, 科技人才信息标准化是人才数据库建设的基础, 人才信息分类编码是数据库建设的基础, 因为科技人才信息分类编码是科技人才信息数据库数据组织、存储、管理和交换的共同基础, 是实现系统数据共享与互操作的重要前提。而我国针对科技人才信息的分类编码的研究工作尚在摸索和实践阶段。通过对我国政府科研管理部门高层次科技人才数据库建设情况调研, 发现44%的科技人才数据库采用《中国学科分类与代码》分类, 12%数据库采用《中国教育部学科分类》, 9%数据库采用《中国职业分类与代码》分类^[2]。

由此可见, 当前科技人才信息没有统一的分类编码方案。大多数数据库是采用现有的国家或部门标准, 标识维度单一, 只能部分揭示信息内容, 不利于科技人才信息

的共享和宏观管理。因此, 建立一套科技人才信息分类编码方案, 对科技人才信息化、标准化建设至关重要。

本文在优化、整合、更新、补充现有的分类编码基础上, 根据科技人才信息的属性特征, 按照一定的原则, 设计一套科技人才信息的分类编码方案, 并应用于2008年中国科学技术信息研究所自建《中国高层次科技人才数据库》中, 数据库收集资源对象为高层次创新型科技人才和潜在科技青年人才。以下内容均以该数据库建设为例进行论述。

2 科技人才信息内容和特点

2.1 科技人才信息内容

科技人才是科学人才和技术人才的略语, 是在社会科学技术劳动中, 以自己较高的创造力、科学的

* 本文为国家软科学研究计划——科技人才信息宏观监测机制研究项目(编号: 2009GXSK047)资助。

探索精神,为科学技术发展和人类进步作出较大贡献的人。科技人才的概念应当大致包含四个要点:具有专门知识和技能;从事科学或技术工作;较高的创造力;对社会作出较大的贡献^[3]。

科技人才信息是指用数字、文字、符号、图片等介质来表征的科技人才的个人基本信息、知识和技能、科技活动、科研成果产出、科研诚信等信息(本文中科技人才信息概念只涉及科技人才个体相关信息,不包含科技人才管理相关信息)。其主要内容如下:

- (1) 个人基本信息:姓名、年龄、性别等;
- (2) 知识和技能信息:学习经历、培训经历等信息;
- (3) 科技活动信息:工作经历、兼职经历、学术活动、评审活动、科研团队等信息;
- (4) 科研成果信息:荣誉奖励、论文、专利、项目、专著等成果关联信息;
- (5) 科研诚信信息:学术道德、科研不端行为、科技信用等信息。

2.2 科技人才信息特点

科技人才信息的主要特点包括:

- (1) 信息范围广:范围涉及经历、成果、学术活动、科研道德、奖励等方方面面;
- (2) 信息类型多:各类信息大多都已分散存储在相关的数据库中,信息类型各异,分类编码体系各异;
- (3) 分类维度多:可以用多维度的属性特征对科技人才进行分类,比如:科技人才的机构特征、地域特征、学科特征、领域特征、群

体关系等。

3 与科技人才信息相关的分类编码

自1979年开始,我国信息分类编码工作发展迅速。国家和各部门都相继制定出台了各类分类编码,其中与科技人才信息相关的分类编码主要内容见表1。

表1 我国与科技人才信息相关的分类编码主要内容

标准级别	分类编码内容
国标	性别(GB/T 2261.1-2003)、国家(GB/T 2659-2000)、中国地区(GB/T 2260-2002)、民族(GB/T 3304-1991)、行业(GB/T 4754-2002)、政治面貌(GB/T 4762-198)、语种(GB/T 4880.1-2005)、职业(GB/T 6565-1999)、学位(GB/T 6864-2003)、职务(GB/T 8561-2001)、奖励类型(GB/T 8563.1-2005)、荣誉类型(GB/T 8563.2-2005)、职业(GB/T 6565-1999)、学科(GB/T 13745-1992)等代码;
教育部标准	中国教育部学科分类、高等学校和科研机构代码
高教行业标准	导师、著作、专利
高等学校标准	技术领域、科技奖、论著收录、项目类型、项目来源
中国图书资料分类法编委会	中国图书资料分类法
国家自然科学基金委员会	中国国家自然科学基金学科分类
科技部	国家科技计划优先发展领域分类

人才信息资源建设的整体需求;

- (2) 内容过时:奖励/荣誉类型、机构、项目类型、项目来源等分类编码亟待更新补充;
- (3) 内容缺失:没有与奖励/荣誉名称、机构单位性质、科研诚信信息相关的分类编码。

4 科技人才信息分类编码原则

4.1 一般原则

- (1) 实用性原则:综合科技人才信息化、标准化建设的需求和目

上述分类编码主要是以行业、学科、图书文献、地域、职业、职务、职称、科研项目、国家科技计划等为分类对象而建立的分类编码,用于科技人才信息分类编码存在以下问题:

- (1) 缺乏顶层规划:当前众多的分类编码系统,只能解决局部需求,缺乏整体设计,不能满足科技

的,使分类便于使用;

- (2) 兼容性原则:尽量按照“国际/国家标准-行业标准-部门标准-更新-自建”的顺序,优先使用现有标准,或在其基础上更新,以便能在其他系统中推广应用;
- (3) 稳定性原则:选择信息最稳定的属性特征作为分类编码的基础和依据;
- (4) 系统性原则:根据概念等级分出上下级类目,避免不同层级和类目之间交叉、重复或遗漏,形成完整的、协调一致的分类体系;
- (5) 可延性原则:预留足够的空码位,以便能增加新的事物和概

念,对下级概念进行延拓和细化;

(6) 灵活性原则:在使用过程中允许在分类编码基础上加前缀码或后缀码,以最大限度地满足需求。

4.2 特殊原则

根据科技人才信息特征,其分类编码还应该遵循以下特殊原则。

(1) 多维度原则:选取多个重要的标志性信息特征进行分类编码,以实现对海量科技人才的定位检索;

(2) 同类合并原则:对科技人才信息中具有相同属性的共性信息集合处理,统一分类编码;

(3) 差异性原则:遵从科技人才信息的差异性,分别制定各种信息分类编码的子体系。

5 科技人才信息分类编码设计

5.1 方案框架设计

以科技人才信息内容为分类编码对象,按照科技人才信息分类编码方案构建原则,建立分类编码方案总体框架(见表2)。

整个分类编码方案由14个分类对象及其相关的26套分类编码组成。

分类编码方案框架最高层是根据科技人才信息属性特征而划分,分为14个对象,分别为个人基本信息、机构、地区代码、学科/学位、技术领域、奖励/荣誉、行业/职业、职务/职称、著作、专利、国内期刊论文、SCI论文、项目、科研诚信。

26套分类编码都是在一个特定的维度表达科技人才信息的某种属性,而分类编码中的每一个分类编码号都包含一个概念,表达一个特

表2 科技人才信息分类编码方案总体框架

分类对象	分类编码内容	功能	标准级别
1.个人基本信息	人的性别代码	辅助标识	国标
	中国各民族名称代码	辅助标识	国标
	政治面貌代码	辅助标识	国标
	语种名称代码	辅助标识	国标
2.机构	高校和科研机构代码	检索维度1	教育部标准+更新补充
	机构单位性质代码	辅助标识	自建
3.地区代码	中国行政地区划分代码	检索维度2	国标
	世界各国和地区代码	辅助标识	国标
4.学科/学位	中国学科分类与代码	检索维度3	国标
	国家自然科学基金学科分类	辅助标识	自科委
	学位代码	辅助标识	国标
5.技术领域	科技计划发展领域分类	检索维度4	科技部
6.奖励/荣誉	奖励/荣誉代码	检索维度5	自建
	奖励/荣誉类型代码	辅助标识	国标
7.行业/职业	国民经济行业代码	辅助标识	国标
	中国职业分类与代码	辅助标识	国标
8.职务/职称	专业技术职务代码	辅助标识	国标
9.著作	著作类别代码	辅助标识	行标
10.专利	国际专利分类号	辅助标识	斯特拉斯堡协定联盟
	专利代码(类别、法律状态、类型、批准形式)	辅助标识	行标
11.国内期刊论文	论文收录代码	辅助标识	高校标
	中图分类号	辅助标识	中图分类法编委会
12.SCI论文	学科类别	辅助标识	SCI
13.项目	项目类型代码	辅助标识	自建
	项目来源代码	辅助标识	自建
14.科研诚信	科研不端行为分类代码	辅助标识	自建

定的内涵和外延。

在26套分类编码中,既有现在正在实行的国家、部门、行业、相关机构的分类编码,也有在其基础上更新补充的分类编码,还有自建的分类编码。

(1) 更新补充:对内容过时的教育部标准《高等学校和科研机构代码》更新补充;

(2) 重新设计:对内容变化比较大的奖励/荣誉、项目类型、项目来源代码进行了重新设计;

(3) 自建:对内容缺失的机构单位性质、科研诚信等信息分类编码进行了自建。

另外,分类编码方案从检索功能的角度,把分类编码划分为2大类:

(1) 检索维度: 选取机构、中国行政地区、学科、技术领域、奖励/荣誉群体这5套分类编码为重点检索维度, 实现对科技人才的定位检索。

(2) 辅助标识: 其他21套分类编码是对科技人才信息的辅助标识。

它们之间相互联系、相互制约、相互依赖、相互补充, 以充分满足科技人才信息资源的规划、数据集标识、信息系统建设等需求。

5.2 分类编码设计

分类编码设计过程主要分为分类编码对象确定、信息分类和信息编码三个部分, 下面以奖励/荣誉信息的分类编码为例, 简述其设计过程。

(1) 对象确定: 分类编码对象是政府和社会对做出自然科学发现、技术发明和改进、科技进步及其他科技成果的科技人才, 依据一定的标准和方式给予的奖励或者

荣誉称号^[4]。

(2) 信息分层: 根据上述分类编码对象的内容, 可以将奖励/荣誉按照授予组织的级别, 分为国家级奖项、行业/部委级奖项、省直辖市级奖项、社会力量奖项、国外重要奖项5大类别。类别根据需求可以扩展。

(3) 信息编码设计: 采用“类别码+地区码+顺序码”的9位分类编码结构。

● 类别码: 前2位英文字母表示类别, PA表示国家级奖项, PB表示行业/部委级奖项, PC表示省直辖市级奖项, PD表示社会力量奖项, PF表示国外重要奖项;

● 地区码: 中间3位阿拉伯数字表示地区。对于中国的奖项选取《中国行政地区划分代码》中的前3位阿拉伯数字; 对于国外奖项选取《世界各国和地区代码》中相应的3位阿拉伯数字;

● 顺序码: 后4位阿拉伯数字为顺序编码, 将同一地区码下的奖项按顺序编码, 一个奖项一个编码,

奖项和编码一一对应。

× ×	× × ×	× × × ×
类别码	地区码	顺序码

6 结语

面向科技人才信息资源的分类编码方案较好地反映了科技人才信息资源的属性特征, 能有效地组织、关联和揭示科技人才信息资源, 充分利用各类的分散信息, 提高信息的有序化程度, 降低数据的冗余度, 基本满足科技人才信息资源信息化、标准化建设的需求。经过在《中国高层次科技人才数据库》建设中的试运行, 反馈效果良好。

同时, 为保证分类编码工作的顺利开展, 我们专门设立了分类编码管理小组, 负责科技人才信息资源分类编码的规划、制定、测试、审核和监督执行等工作。分类编码一旦制定, 即应成为各部门共同遵守的规范性文件, 不能随意改动。

参考文献

- [1] 国家统计局. 中国科技统计年鉴(2009)[M]. 北京: 中国统计出版社, 2010.
- [2] 国家软科学研究计划项目(项目编号: 2008GXS4K062). 科技人才信息资源整合机制研究[R]. 2010.
- [3] 刘茂才, 段成瑞, 王青, 张纪明. 人才学辞典[Z]. 成都: 四川省社会科学院出版社, 1987.
- [4] 宁国栋, 张涛伟. 试论我国的科技奖励法律制度[J]. 技术与创新管理, 2008(3).

作者简介

吴晓莉 (1961-), 女, 中国科学技术信息研究所副研究馆员, 硕士, 主要研究方向: 科技人才. E-mail: wuxiaoli611117@126.com

A Classification and Coding Solution for Information Resources of Science and Technology Talents

Wu Xiaoli, Wang Yunhong / Institute of Scientific and Technical Information of China, Beijing, 100038

Abstract: This paper proposed a classification and codification solution to S&T talents information in order to optimize the existing classification and codification system. The solution includes the property characteristics of resources including talents' basic information, knowledge and competence, scientific activity, production as well as their credibility information. The solution could organize, relate and reveal science and technology talents information, which would be very important to the establishment of database and data sharing.

Keywords: Classification, Coding, S&T talents, Information resource of talents

(收稿日期: 2012-04-23)