

医药卫生科学数据分类与编码的研究

张黎黎¹, 刘丽华^{2a}, 徐勇勇¹, 尹岭^{2b}

[摘要] 目的 研究医药卫生科学数据的分类与编码方法,以保证科学数据得到有效的管理。方法 采用线分类法,建立医药卫生科学数据分类框架;使用组配技术进行编码,实现体系分类法的组配化,建立医药卫生科学数据的多轴分类集合。结果 建立了医药卫生科学数据分类框架;确立了医药卫生科学数据分类框架类目内容;设计了医药卫生科学数据编码方案。结论 分类编码研究实现了针对数据的归并和信息的组织。

[关键词] 医药卫生;科学数据;线分类;编码;组配

Classification and Coding of Medical Science Data ZHANG Li-li, LIU Li-hua, YIN ling, et al. Department of Health Statistics, the Fourth Military Medical University, Xi'an 710032, Shaanxi, China

[Abstract] **Objective** To develop the classification and coding of medical science data so as to merge and manage the data effectively. **Methods** Classification frame was developed with linear classification, coding with matching design. **Results** The medical science data classification frame, the contents of the frame and their code have been developed. **Conclusion** The development of the classification and coding can merge and organize the medical science data effectively.

[Key words] medicine; science data; linear classification; code; match

中图分类号:R197.324 文献标识码:A 文章编号:1006-9771(2006)05-0452-02

[本文著录格式] 张黎黎,刘丽华,徐勇勇,等.医药卫生科学数据分类与编码的研究[J].中国康复理论与实践,2006,12(5):452-453.

医药卫生领域已经积累了大量的科学数据,但是这些科学数据资源缺乏系统化的归纳、层次化的标识和规范化的表达,缺乏统一的、可操作的资源调查、规划、设计与描述,导致科学数据资源使用效率低,重复投入,浪费严重。因此,建立医药卫生科学数据分类与编码方案,有重要的现实意义。

医药卫生科学数据分类与编码研究的最小单元是数据集。分类与编码是将具有某种共同特征的数据集归并在一起,使之与不具有上述共性的数据集区分开来,然后设定某种符号体系进行编码,以保证科学数据得到有效的管理。

1 分类与编码的方法

采用线分类法,建立医药卫生科学数据分类框架;使用组配技术进行编码,实现体系分类法的组配化,建立医药卫生科学数据的多轴分类集合。

1.1 线分类法 线分类法也称“层次分类法”,它是将初始的分类对象按选定的属性作为划分基础,逐渐分成相应的若干个层次级类目,并排列成一个有层次的、逐级展开的分类体系。表现形式是大类、中类、小类等。各类之间构成并列或隶属关系,既不重复也不交叉^[1]。

1.2 组配 组配是分类的类目之间以及与通用性表进行概念的匹配,即在标引和检索时,通过组配方法实现医药卫生科学数据内容的揭示,通过类号与类号的逻辑组合来表达和描述医药卫生科学数据内容主题与概念的技术措施。采用组配方法有利于增强数据聚类的灵活性,建立医药卫生科学数据的多轴分类集合。

2 医药卫生科学数据分类与编码

作者单位:1. 第四军医大学卫生统计学教研室,陕西西安市710032;2. 解放军总医院,a. 医疗统计科;b. 神经信息中心,北京市100853。作者简介:张黎黎(1981-),女,山东东营市人,硕士,主要研究方向:医院信息资源分析利用、医学信息标准化。通讯作者:刘丽华。

科学数据分类编码中的每个类目都是一个特定的主题,表达一定学科知识的内涵和外延。医药卫生科学数据分类与编码通过科学分类体系序列类目,来容纳大量的科学数据集,从而成为管理数据集的工具。

2.1 分类框架 医药卫生科学数据分类框架由主分类表和复分类表两部分组成,主分类表、复分类表分别面向该学科数据集的不同主题采用线分类法进行类目设置。

2.1.1 主分类表 主分类表展开的层次决定着分类导航系统的导航程度,层次越多越深,知识被组织得越细密,用户查找信息耗费的时间和精力就越多。根据目前网络信息的实际情况和用户的浏览习惯,参考诸多信息分类与编码研究的文献报道,分类框架主分类表的层次应控制在3~6级^[2]。据此,将医药卫生科学数据分类框架主分类表的层次设置为亚门类、大类、小类。

2.1.2 复分类表 主分类表中有许多类目由于进一步细分时要采用统一标准,为了增加标引数据集的能力,缩小分类表的篇幅和细分方便,特别设立了复分类表。复分类表的设立应该考虑医药卫生科学数据的概念内容和学科等综合因素进行设置。本分类框架中拟采用的复分类表包括地区表(国际采用世界地区表,国内采用GB 2260-90《中华人民共和国行政区划代码》)、全国组织机构代码、疾病和有关健康问题的国际统计分类——第10次修订本(ICD-10)、GB/T 13745-92学科分类与代码表。

2.2 医药卫生科学数据分类框架类目内容

2.2.1 主分类表类目内容 医药卫生科学数据主分类表最高层的分类遵循学科分类方法,分为基础医学、临床医学、公共卫生、中医学、药学和特种医学6个亚门类。亚门类下位类的划分遵循分类基本原则,类目的确定以科学内容主题、专业领域内习惯和规范化3个主要因素为依据,对类目概念的划分争取实现与各专业领域分类类目的参照和对应。主分类表类目内容及其编码见表1。

表 1 医药卫生科学数据主分类表亚门类、大类、小类及其编码

亚门类		大类及小类	亚门类		大类及小类
11 基础医学	11 医学生物学	11 病原生物学,12 医学生物化学,13 医学遗传学,14 医学实验动物学	12 中医	医药教育,15 中医药科研,16 中医药经济	
	12 人体机能	11 人体生理信息,12 医学免疫信息,13 病理生理信息,14 病理学信息,15 药理信息,16 生物医学工程信息		11 中医基础理论,12 中医诊断学,13 中医治疗学,14 中医病证,15 中医保健护理康复,16 少数民族医学	
	13 医学相关学科	11 流行病学,12 医学统计学,13 生命伦理学		13 中药	11 中药材,12 单味药,13 中药饮片及其炮制,14 中药化学,15 中药药理学,16 中药毒理学,17 剂型,18 复方,19 中成药
	14 人体解剖学	11 大体解剖学,12 神经解剖学,13 影像解剖学,14 组织学,15 胚胎与发育生物学,16 细胞生物学		14 针灸	11 经络腧穴,12 针灸仪器设备,13 针刺疗法,14 灸法,15 穴位疗法(经络疗法),16 针灸疗法应用
12 临床医学	11 临床诊断信息	11 症状诊断信息,12 物理诊断信息,13 电诊断信息,14 鉴别诊断学,15 机能诊断学,16 预后及劳动鉴定	15 药 学	15 古代文献	11 医经,12 基础理论,13 伤寒金匱,14 诊法,15 针灸按摩,16 本草,17 方书,18 临证各科,19 养生,21 医案医话医论,22 医史,23 综合
	12 临床治疗信息	11 药物与化学治疗,12 手术治疗,13 物理与自然治疗,14 生物治疗,15 介入治疗,16 血液治疗,17 组织治疗,18 特殊治疗		11 药品监督管理	11 法律法规,12 规范与标准,13 产品管理,14 机构与人员管理
	13 临床影像信息	11 超声波诊断,12 核磁共振成像,13 电子计算机体层扫描,14 放射线诊断,15 放射性核素成像,16 发射型计算机体层显像,17 热成像,18 其他		12 药物资源	11 数据信息,12 标本信息
	14 临床实验室信息	11 血液学检验,12 尿液检验,13 粪便检验,14 脑脊液检验,15 微生物学检验,16 免疫学检验,17 组织学检验,18 其他检验信息		13 药物研发	11 图谱信息,12 化学药品研发,13 中药研发,14 生物制品研发
13 公共卫生	15 临床护理信息	11 基础护理信息,12 一般护理信息,13 专科护理信息,14 特殊护理信息,15 临终关怀	16 特种医学	14 药品生产	11 化学药品生产,12 中药生产,13 生物制品生产,14 制药中间体及药用辅料
	16 临床康复医学信息	11 社会康复,12 医学康复,13 教育康复,14 康复医学工程		15 药品经营	11 药品批发,12 药品零售,13 招标采购,14 第三方物流配送,15 医院用药分析,16 OTC 市场
	17 医院管理信息	11 政策法规,12 卫生经济,13 仪器设备,14 临床药品管理,15 循证医学,16 卫生资源		16 药品使用	11 医疗,12 自我诊疗
	11 疾病预防控制信息	11 监测信息,12 干预信息,13 通告信息		17 医药文献与成果转让	11 医药文献,12 成果转让
14 中医学	12 公共卫生服务类信息	11 卫生检测,12 预防与防护,13 健康促进	18 医药专利	11 发明专利,12 实用新型专利,13 外观设计专利	
	13 公共卫生管理类信息	11 政策法规,12 专题调查,13 综合评价,14 卫生资源,15 科研管理		11 特殊损伤医学防护信息	11 核辐射损伤医学防护信息,12 化学中毒医学防护信息,13 生物危害医学防护信息,14 火器伤医学防护信息,15 其他特殊损伤医学防护信息
	14 公共卫生监督类信息	11 执法检查,12 产品审批(鉴定),13 行政处罚,14 行政复议,15 案例		12 特殊环境作业医学信息	11 航海医学信息,12 潜水医学信息,13 航空医学信息,14 航天医学信息,15 特殊地理气候医学信息
	11 中医药事业	11 中医药管理,12 中医药发展,13 机构及人员,14 中		13 特殊职业医学信息	11 运动医学信息,12 法医学信息,13 军队疾病预防控制信息

2.2.2 复分类表类目内容 本分类框架采用的复分类表将等效采用所列类表的既定分类及代码内容。其主要目的在于充分利用复分、组配方法,确定科学数据的分面组配内容,使得许多类目再进一步细分时采用同一标准,增加标引数据的能力、缩小类目表篇幅,使分类具有最大的容纳性和灵活性,相对应地确立分面和顺序制相结合的新型编码设计,从而实现数据集的唯一标识^[3]。

2.3 医药卫生科学数据编码设计 按照《标准化信息导则:信息分类编码的基本原则和方法》规定,信息的编码应遵循惟一性、合理性、可扩充、简化、适用和规范化原则。编码的类型分数字型、字母型和混合型 3 种,一般认为数字型编码较好^[4],因此本文主分类表编码采用数字码。

医药卫生科学数据集采用分段编码设计,码位分为 6 段,分别为主分类码、地区码、组织机构码、疾病码、学科分类码。其完整的编码由 5 段 34 个数字组成,形式为 ××××××-××××××-××××××××××-××××.××××-×××.××××。其中,第 1 段 6 个数字表示主分类码,即亚门类、大类、小类编码分别设为 2 位,由 11~99 组成,编码不使用 20、30 等个位为 0 的两位整数;第 2 段 6 个数字表示地区码(如,北

京市 110000;北京市海淀区 110108);第 3 段 9 个数字表示组织机构码(如,中国医学科学院 400012916);第 4 段 6 位表示疾病码(如,伤寒 A01.001);第 5 段 7 个数字表示学科分类码(如,内科学 320.2400;心血管病学 320.2410)。数据集名称不体现某个复分表分类属性时,整段编码用 0 表示。

3 结论

医药卫生科学数据涉及基础医学、临床医学、公共卫生、中医学、药理学及特种医学多方面的知识,内容十分庞杂。本文仅仅提出了医药卫生科学数据分类的一种构想,能否全面涵盖领域数据信息,代码是否具有唯一性还有待于进一步检验,每种分类或类目下具体的指标有待于进一步细化。同时建议在医药卫生科学数据资源的建设中,要加强科学数据分类的研究,提高医药卫生科学数据资源的管理和共享程度。

[参考文献]

[1]刘植婷.信息分类编码标准化综述[J].中华医学图书馆杂志,2000,5:8-10.
[2]董琳.网络信息分类组织的发展趋势与标准化[J].图书情报知识,2004,2:65-67.
[3]中国图书馆图书分类法编委会.中国图书馆图书分类法[M].北京:书目文献出版社,1990.
[4]郭书普.网络农业信息分类和编码的研究[J].农业图书馆学报,2003,6:139-141. (收稿日期:2006-01-23)