

组配法在医药卫生科学数据分类编码中的应用

张黎黎¹, 刘丽华^{2a}, 徐勇勇¹, 尹岭^{2b}

[摘要] 目的 探讨组配方法在医药卫生科学数据分类与编码中的应用, 建立医药卫生科学数据的多轴分类集合, 实现科学数据的规范表达和有效管理。方法 采用组配方法, 对医药卫生科学数据设计分段编码方案。结果 主分类表设计 5 个码位, 表示三个层级, 即门类(第一位)字母码、亚门类(第二、三位)数字码和大类(第四、五位)数字码; 复分类表等同采用既定分类方法中的分类编码体系。结论 组配法与传统分类法相结合可实现分类法等级列举式向分面组配式发展, 从而达到对医药卫生科学数据归并和信息组织的目的。

[关键词] 医药卫生; 科学数据; 线分类; 组配

Application of Matching in Medical Scientific Data Classification and Coding ZHANG Li-li, LIU Li-hua, XU Yong-yong, et al. The Department of Health Statistics, the Fourth Military Medical University, Xi'an 710032, Shaanxi, China

[Abstract] Objective To explore the application of matching in medical scientific data classification and coding and establish the multi-classific groups and manage the datas effectively. Methods Medical science data subsection coding rules were designed using matching method. Results Main classification had five codes, which stood for three grades. Phylum (the first grade) was letter code, subphylum (the second and third grades) was number code, great class (the fourth and fifth grades) was number code. In the process of designing matching classification, given classification and coding system were used. Conclusion The merging between traditional classification and matching method is very important in the process of specializing classification to multi-factor matching, so as to merge and organize the data information.

[Key words] medicine and health; scientific data; linear classification; match

中图分类号: R195.1 文献标识码: A 文章编号: 1006-9771(2006)04-0364-02

[本文著录格式] 张黎黎, 刘丽华, 徐勇勇, 等. 组配法在医药卫生科学数据分类编码中的应用[J]. 中国康复理论与实践, 2006, 12(4): 364-365.

近年来, 随着科学技术的迅猛发展, 出现了知识爆炸、信息资源剧增的局面, 因此, 信息的组织、知识的表示、提高信息利用的效率成为非常重要和迫切的研究课题。

组配是分类的类目之间以及与通用性表进行概念的组配。所谓组配, 就是在标引和检索时, 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。采用组配方法的目的在于增强数据聚类的灵活性、增加数据检索的入口, 并取得检索系统轮排的效能, 使体系分类法适应网络环境下机检的需要。本分类研究旨在探讨组配法在医药卫生科学数据分类编码中的应用。

1 资料与方法

医药卫生科学数据分类体系包括主分类表和复分类表。主分类表及复分类表各自采用线分类法围绕数据资源的主题展开等级列举, 从而实现围绕数据资源不同属性的类目细分。复分类表的设立应考虑医药卫生科学数据的内容概念和学科等综合因素进行设置。本分类研究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, 采用组配方法, 对医药卫生科学数据设计分段编码方案。

2 结果

设计的医药卫生科学数据分段编码方案见图 1。本数据分段组配编码能够清晰地表达概念主题的含义, 充分利用数字型、字母型和混合型等不同编码类型^[1], 主分类表设计 5 个码

位, 表示三个层级, 即门类(第一位)字母码、亚门类(第二、三位)数字码和大类(第四、五位)数字码; 复分类表等同采用既定分类方法中的分类编码体系。

本编码方案应用组配法的要点是: ①医药卫生科学数据分类时类目之间的组配是将概念相关的两个或多个类号通过组配符号连接在一起, 以表达一个分类表中没有列举的比较复杂或更专指的主题; ②凡是主表中已经存在的主题类目, 则将该科学数据直接归入该类, 均不再应用组配方法表达主题概念; ③科学数据内容的主题之间具有并列关系, 比较关系时直接用相关类号组配; ④科学数据内容的主题之间具有因果关系、作用与被作用关系、应用与被应用关系、影响与被影响关系, 且主题的类目下有该表示关系的类目, 直接应用类目进行组配对科学数据标引; ⑤当希望通过利用组配技术实现科学数据内容新主题的标引时, 组配的类目要选择最细分的类目级别; ⑥如果主表中类目下有关于主题归类划分的注释, 应该首先按照注释的要求进行分类标引, 不能随意使用组配技术。

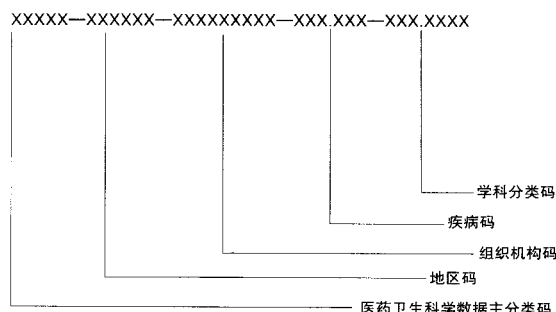


图 1 医药卫生科学数据分段编码方案

3 讨论

分类是人类认识事物的基础, 科学数据分类编码的目的在于准确地识别数据, 对数据实施有效的管理, 并能按类别开发利用数据。在高层次上进行数据的归并和信息的组织, 对科学

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, 山东广饶县人, 硕士, 主要研究方向: 医院信息资源统计利用, 卫生信息标准化。通讯作者: 刘丽华。

数据进行分类编码,将具有某种共同特征的数据归并在一起,使之与不具有上述共性的数据区分开来,然后设定某种符号体系进行编码,使之能够进行计算机或人工识别和处理,可保证科学数据得到有效的管理,并能支持高效率的查询服务。

长期以来,医药卫生领域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、可操作的资源组织与描述,科学数据资源杂乱无章,数据利用效率极低。医药卫生科学数据分类与编码的研究就是将具有某种共同特征的数据归并在一起,然后设定某种符号体系进行编码,保证科学数据得到有效的管理。

目前,传统分类法的研究主要集中在:①以学科分类为基础的图书馆分类法;②实现了分类语言与主题语言融合的标杆公司信息资源分类体系;③主题聚类为主,学科聚类为辅的主题分类系统(如 yahoo 网站、sohu 网站)等^[2]。

传统的分类法主要是等级列举式,这种方法通过构建层级展开的树型结构仅能达到类目细分的目的,如《中国图书资料分类法》前三版的类目表构建。等级列举式树型结构的局限性在于既不可能把古往今来的一切主题都囊括无遗,也不可能把主题之间的确存在的各种关系全部揭示出来,更不能适应科学技术的飞跃发展^[3]。

组织科学数据的关键在于如何将这些数据资源进行系统化归纳、层次化标识和规范化表达,对于医药卫生科学数据而言,就是如何将数据内在的主题充分表示,揭示数据之间的确存在的各种关系,以适应科学数据主题概念的不断扩展。

在医药卫生科学数据的分类过程中,组配技术是利用分类表中已经存在的表达简单主体概念的类号,按照一定的规则组合成一个复合类号,用来表达分类表中没有的复杂的概念,对

科学数据进行标引。利用组配技术可以提高描述科学数据内容的能力,提高科学数据类目的细分程度,规范类目共性区分标准和组合类号的规律性。

组配有 3 种方式,即联合组配、关联组配和复分组配。凡医药卫生科学数据内容的主题概念涉及两个以上类目,而且它们之间是一种并列或交替的关系时,应采用联合组配(用“+”号);凡医药卫生科学数据内容的主题概念涉及两个以上类目而且它们之间是一种交叉、细分、应用、因果和影响关系时,应采用关联组配(用“:”号);凡医药卫生科学数据内容的主题概念涉及主类号需用通用复分类表分类时,应采用复分组配(用通用复分表中的几种标识符号)。

医药卫生科学数据分类编码普遍采用组配技术,力求克服传统等级列举式分类法的不足。本编码方案借鉴《中国图书资料分类法(第四版)》的复分、组配方法,确定医药卫生科学数据的分面组配原则,许多类目进一步细分时要采用同一标准,实现传统的等级列举式向分面组配式发展,增加标引数据的能力,缩小小类目表篇幅,方便细分,而相对应的确立分面和顺序制相结合的新型标记制度,使分类具有最大的容纳性和灵活性,实现数据集的惟一标识。

[参考文献]

- [1]郭书普.网络农业信息分类和编码的研究[J].农业图书情报学刊,2003,6:139-141.
- [2]刘植婷.信息分类编码标准化综述[J].中华医学图书馆杂志,2000,5:810.
- [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.