

医学分类主题一体化系统建设^{*}

李军莲 李丹亚 阮学平 钱 庆 金 晶

(中国医学科学院医学信息研究所 北京 100020)

〔摘要〕 为适应医学学科发展,更好地发挥医学分类主题一体化知识组织工具在医学文献组织与利用方面的作用,基于 1992 年编制出版的《中国图书馆图书分类法(R 类)与医学主题词表(MeSH)中医药学主题词表对应表》(以下简称:《医学对应表》),初步建立计算机化、可持续发展的医学分类主题一体化系统,实现《医学对应表》的动态更新与及时发布。

〔关键词〕 医学;分类主题一体化;中国图书馆分类法;医学主题词表

Construction of Medical Classification Subject Entirety System LI Jun-lian, LI Dan-ya, RUAN Xue-ping, QIAN Qing, JIN Jing, Institute of Medical Information, Chinese Academy of Medical Sciences, Beijing 100020, China

〔Abstract〕 In order to adapt to the development of medical sciences and make the classification subject entirety knowledge organization tool play a better role in medical literature organization and utilization, based on *Chinese Library Classification (R type)* and *Medical Subject Headings (MeSH) TCM Thesaurus Correspondence Table* (referred to as: *Medical Correspondence Table*) compiled and published in 1992, the paper primarily constructs the computerized and sustainable developing medical classification subject entirety system, which can be used to realize the dynamically updating and timely releasing of *Medical Correspondence Table*.

〔Keywords〕 Medicine; Classification subject entirety; Chinese Library Classification; Medical subject headings

1 引言

分类主题一体化系统又称分类主题一体化语言,或分类主题一体化词表,具体指在一个知识组织系统中通过对分类表部分和叙词表部分的术语、参照、标识及索引实施统一控制,使知识组织内容揭示能够有机融合,发挥优势互补效应^[1]。分类主

题一体化的研究最早始于 20 世纪 60 年代,至今已经历了必要性研究、可行性研究和实践研究 3 个阶段,理论研究日趋成熟,实践方法日趋多样^[2]。在以本体为代表的新型知识组织系统实践研究推进相对缓慢的今天,分类主题一体化依然是情报检索语言发展的主要方向之一,尤其是在实践应用研究方面。这不仅是新时期文献编目、标引及检索的需要,也是深入开展知识服务的需要。

1992 年,中国医学科学院医学信息研究所编制出版了《中国图书馆图书分类法(R 类)与医学主题词表(MeSH)中医药学主题词表对应表》^[3](以下简称《医学对应表》),将《中图法》第 3 版 R 类与 MeSH 1990 年版、《中国中医药学主题词表》1987 年版相对应,是我国医学领域第 1 部最具代表性和有影响力的分类主题一体化知识组织工具,在

〔收稿日期〕 2010-12-13

〔作者简介〕 李军莲,副研究馆员,发表论文 10 余篇,参编论著 5 部。

〔基金项目〕 中国医学科学院医学信息研究所中央级公益性科研院所基本科研业务费专项“医学分类主题一体化系统建设”(项目编号:07R011)。

推动医学文献组织利用方面发挥了重要的作用。

10 余年来,随着医学学科的发展及源词表的不断修订,《医学对应表》虽也进行了不同程度的更新,但相对而言内容较为陈旧,已远远落后于分类表、主题词表的发展变化,在揭示医学新学科、新主题方面,在对医学信息的组织处理方面已经不能满足现实的需要。在此背景下,研制能够适应医学科学发展,实现《医学对应表》及时更新与发布的计算机化的可持续发展的医学分类主题一体化系统迫在眉睫。

2 医学分类主题一体化系统建设规划

2.1 建设目标

建立一个计算机化的可持续发展的医学分类主题一体化系统。在对 1992 年版《医学对应表》的内容和规范进行全面修订和发展的基础上,初步建立医学分类主题一体化词表的建设、维护与发布平台,实现《医学对应表》内容的计算机辅助建设与动态更新,及时揭示医学新学科、新主题,满足文献分类主题标引和图书馆编目人员主题编目时多途径、多方式实现主题与分类间的相互查询、快速浏览及多向链接的需要。

2.2 建设原则

医学分类主题一体化系统是集研究与实际应用为一体的医学知识组织系统,在其建设中重点遵循如下原则:(1) 实用性原则。满足现阶段《医学对应表》建设、发布及动态维护的实际需求。(2) 易用性原则。提供简单、方便、友好的用户界面,易于理解、学习和操作,使用户能够快速地掌握系统使用。(3) 标准化原则。系统在设计和实现时,尽量采用业内标准或参照同行标准制订相应的标准规范,以保证本系统数据及应用的标准化。(4) 兼容性原则。系统尽量保证与已有相关系统(如 CBM、WBM)的兼容性,符合已有用户的操作习惯,同时提升用户的使用体验。(5) 开放性原则。提供开放的、易于维护的、灵活的、易于扩展的、统一的标准

数据接口,全面支持 XML、SOAP、Web Service 等当前业界通用的开放标准,保证系统能够与其他平台的应用系统、数据库等相互交换数据并进行应用级的互操作性和互连性。(6) 安全性原则。系统能够根据工作任务的不同设定不同的用户权限,尤其是在数据编辑、审校等方面;同时针对数据本身采取多种备份方式保证数据的安全提供,全面确保系统数据的安全稳定。

2.3 类型定位

按照组织结构的不同,分类主题一体化词表可分为分面叙词表、叙词表式索引、分类表——叙词表双向对照索引和集成叙词表 4 种类型^[1]。分类表——叙词表对照索引是目前存在的分类主题一体化词表中较常见的类型,1992 年版的《医学对应表》便属于此种类型,医学分类主题一体化系统是其延续。所谓分类表——叙词表双向对照索引,又称分类叙词对应表,即为一部分类表与一部叙词表编制类目与叙词对应关系的索引,具体表现为在每个类目(类号)后列出相对应的叙词或叙词的组配形式,而在每个叙词或叙词串后列出相对应的一个或多个类号,是分类检索语言与主题检索语言兼容、互换的有效工具。

2.4 系统结构框架

从整体结构来看,医学分类主题一体化系统主要包括 3 个层面,分别为标准规范层、词表建设维护层和服务应用层,其总体框架,见图 1。

标准规范层主要是定义分类主题对应规则、有关的技术标准和元数据结构,是《医学分类主题对应表》建设维护与发布应用的前提。词表建设维护层以《医学分类主题对应表》的建立、更新维护为核心,同时包括对《中图法》R 类、MeSH 中文版、《中国中医药学主题词表》各独立分类表、主题表的更新维护。在服务应用层,本项目仅实现词表查询服务系统的设计开发,词表在数据库检索系统、计算机辅助标引系统中的应用通过接口调用来实现。

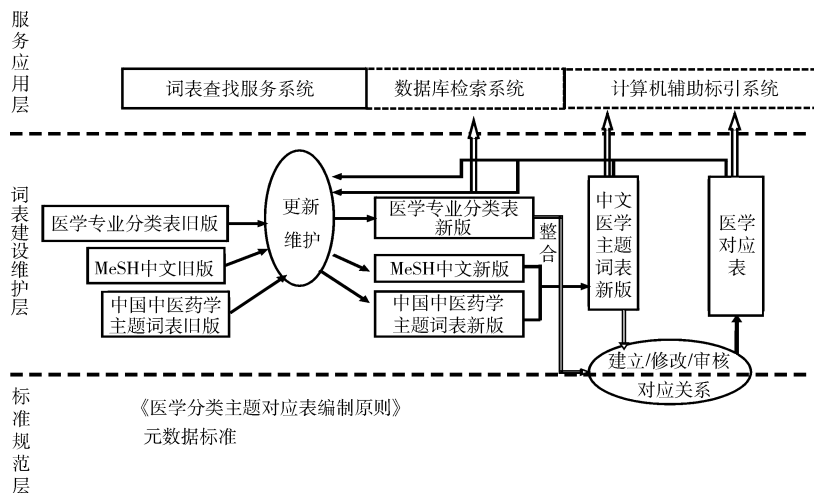


图1 医学分类主题一体化系统总体框架

3 词汇对应规则和元数据方案

从分类表——叙词表双向对照索引建立的可行性及其功能发挥的有效性来看，除了良好的词汇对应基础，合理的对应（映射）规则及完整的元数据结构设计也是研制有实际应用价值的现代分类表——叙词表双向对照索引的关键。

3.1 词汇对应规则

在分类标识和主题标识之间建立正确的对应（映射）关系是分类表——叙词表对照索引的结构基础，也是实现分类标识和主题标识之间相互转换的关键环节。在充分比较分析 1992 年版《医学对应表》编制原则、“医学主题词——分类号对照表”主题分类对应原则的基础上，结合用户反馈，对 1992 年版《医学对应表》编制原则进行了全面修订和补充。

3.1.1 一般类目主题词对应规则 在追求概念映射目标指引下^[4]，分别从对应主题词和类目的优先选择、类目释义的揭示、唯一性、排他性、专指性等方面，制定分类标识和主题标识之间对应（映射）关系。

关于专指性：对应时，必须使用主题词表中的专指主题词。当无专指度相当的主题词时，可依次选用主题词与副主题词组配、主题词与主题词组配进行对应，组配主题词数量以 2 个为宜，最多不能

超过 3 个；无法组配对应时，用上位类目对应的主题词对应。组配主题词之间以“|”分隔。

示例 1：R541.3 高血压性心脏病 | 高血压/并发症 | 心脏病/病因学

示例 2：R543.15 主动脉扩张 | 扩张，病理性 | 主动脉疾病

示例 3：R573.31 浅表性胃炎 | 胃炎

关于优先选择：R2 中国医学部分首选与《中国中医药学主题词表》对应。《中国中医药学主题词表》中确无对应主题词时，可考虑用 MeSH 中相应主题词对应。

示例：R272.22 麻疹 | 麻疹△

R511.1 麻疹 | 麻疹

R254.8 鼠疫 | 中医内科疾病 | 鼠疫

关于无法匹配的类目或主题词：MeSH、《中国中医药学主题词表》中的某些主题词若无确切类目可对应时，采用上位归类或靠类的方法对应应在相关类目之下；对具有多重属性可归入两个或两个以上类的主题词，选择一个类作为主要类，其余类对应时均加“< >”符号，以示区别，最多给出 3 个学科类号。

3.1.2 特殊类目主题词对应规则 主要是为解决分类表中的类组类目（如 R12 环境卫生、环境医学）、“其他类”类目、“一般性问题”和“各种××”、“按××分”、“各国××”等特殊类目、仿分类目、交替类目、通用复分表类目和专业复分表类目与主题词之间的映射问题。

关于“各种××”、“按××分”、“各国××”之类的特殊类名类目的对应：对应时采用正式主题词。若产生歧义则采用上位主题词。

示例：R818. 052 各种制剂疗法 | 辐射损伤/药物疗法
R933/937 各国药材分布、药材志 | 地道药材 |
地理位置

关于交替类目的对应：对应时一律将对应主题词（包括需归类对应的主题词）加方括号“[]”，不加参见注释。原则上，R 类与 R 类之外的类目交替时，启用 R 类中的相应类目为正式类目，此时其对应的主题词（包括需归类对应的主题词）不再加方括号“[]”。

示例：R34 人体生物化学、分子生物学 | 生物化学；
分子生物学

关于专业复分表类目的对应：R 类共有三个专类复分表，依次为：中医临床专类复分表、药材专类复分表、临床专类复分表，分别与相应的副主题词直接对应。

3.2 元数据方案

元数据是资源描述、存储、表达、转换、集成和检索等的基础。医学分类主题一体化词表作为现代意义上的知识组织工具，必须建立一套完整的元数据标准。就当前的数据基础来看，用于构建医学分类主题一体化词表的分类表和主题词表在元数据结构上差异较大。经认真分析和反复讨论，结合医学分类主题一体化词表建设的实际需要和现实基础，在参照 MARC21 规范数据格式的基础上初步确定医学分类主题一体化词表由两个子表组成，分别为《分类主题对应表》和《主题分类对应表》，均独立于《分类表》和《主题词表》，具体的元数据结构，见表 1。

表 1 《分类主题对应表》元数据结构片段

字段名	字段含义	字段名	字段含义
ui	分类 UI 号	ui	主题 UI 号
lh	分类号	zpbj	组配标记
lm	分类名	namezh	主题词
yxzt	优先主题	lh	分类号
ptzt	普通主题	lm	类名
...	...	...	...

4 建设维护系统

4.1 《医学对应表》内容建设机制

为保障分类主题间对应关系的准确性与一致性，结合当前知识组织系统自动映射技术发展、应用现状^[5]，决定采用计算机辅助人工方式建立分类主题间的对应关系，结合计算机辅助审核校验、数据转换全面构建《医学对应表》，着力解决概念间对应关系中的双向性、一对多与多对一等问题，具体的实现步骤，见图 2。

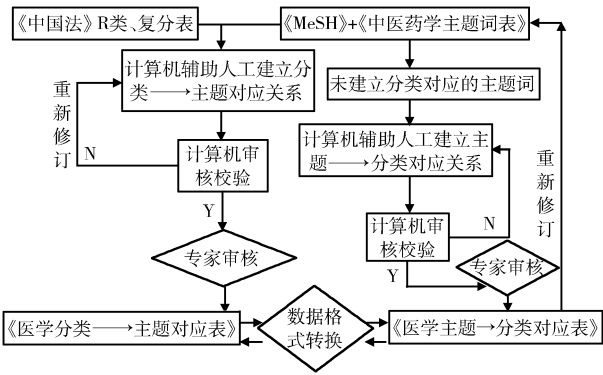


图 2 实现步骤

4.2 系统功能结构

《医学分类主题对应表》建设维护系统以《医学对应表》为主体，以《中国图书馆分类法》R 类、《中文医学主题词表》（CMeSH，含《医学主题词表》（MeSH）、《中医药学主题词表》）为基础。该建设维护系统由多个可扩展的子系统构成，能够提供比较完善的词表系统维护管理、词表维护管理功能，可同时挂接多个词表系统，用户可以根据自己的需求修改词表结构。图 3 是《医学分类主题对应表》建设维护系统的功能体系结构图。主要由《医学分类主题对应表》编制子系统〔主体部分〕、分类表维护管理子系统〔辅助部分〕和主题词表维护管理子系统〔辅助部分〕构成。

4.2.1 《医学分类主题对应表》编制子系统 用以完成《医学对应表》内容的加工，实现实时提交、更新和保存，提供保存前自动校验和数据安全

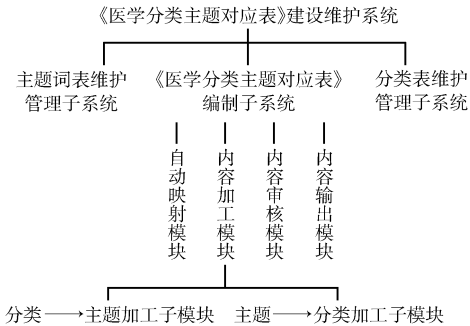


图3 建设维护系统功能体系结构

保障等功能，主要由自动映射模块、内容加工模块、内容审核模块和输出功能模块组成。其中，内容加工模块包括“分类——主题加工”和“主题——分类加工”两个子模块，分别用以提供从类目和主题入手，建立分类与主题之间的对应关系。

4.2.2 分类表维护管理子系统 分类表维护管理子系统的基本功能是支持已有分类类目及其所有数据项的导入、导出；实现对类目数据的增删改、简单的逻辑校验、查询浏览、数据统计分析、打印输出和系统维护等功能。

4.2.3 主题词表维护管理子系统 其基本功能是支持已有主题词及其所有数据项的导入、导出；实现对主题词数据的增删改、简单的逻辑校验、查询浏览、数据统计分析、打印输出和系统维护等功能。

5 发布系统

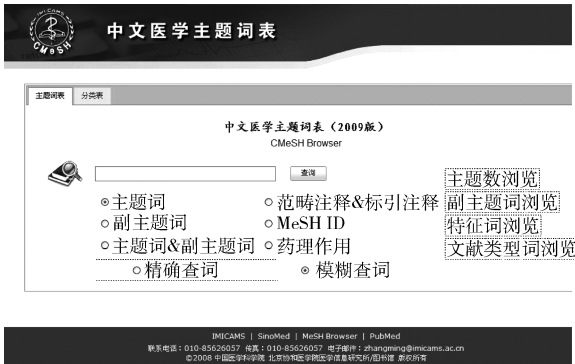


图4 词表发布系统首页

词表发布系统在继承现有中文医学主题词表检索系统^[6]部分功能的基础上，集成医学主题词注释表、医学主题词树形结构表、《中国图书馆分类法》R类、《医学分类主题对应表》的所有内容，具有主题、分类多途径检索浏览功能，见图4所示。

6 结语

本研究在相关标准规范和内容建设机制的研究制定基础上，初步建立了一个计算机化的可持续发展的医学分类主题一体化系统，在很大程度上解决了传统医学知识组织工具无法及时揭示医学新学科、新主题的缺陷，满足了医学信息组织处理现实需要。今后，随着国内数字图书馆知识组织与知识服务研究理念的不断深入，将围绕现阶段知识组织工具的发展特点继续开展生物医学领域知识组织系统的研制、建设及应用研究，研究重点将逐步由单一知识组织系统的独立建设转向多类型知识组织系统整合发展、从各学科的局部建设转向跨领域集成发展、从单一语种的构建向多语种对照和映射发展、从简单/线性语义关系建立向复杂语义网发展。

参考文献

1 杨琼. 分类主题一体化——《中国分类主题词表》发展的方向 [J]. 高校图书情报论坛, 2003, (4): 48 - 49, 30.

2 王丽. 浅谈分类主题一体化的发展 [J]. 佳木斯大学社会科学学报, 2002, (4): 144.

3 林美兰. 中国图书馆图书分类法 (R 类) 与医学主题词表 (MeSH) 中医药学主题词表对应表 [M]. 北京: 中国科学技术出版社, 1992.

4 阮学平. 关于医学分类表与主题词表再对应的思考 [J]. 医学情报工作, 2002, (3): 181 - 183.

5 司莉. 知识组织系统的互操作及实现 [J]. 现代图书情报技术, 2007, (4): 29 - 34.

6 李丹亚, 胡铁军, 诸文雁, 等. 中文医学主题词表检索系统 [J]. 中华医学图书馆杂志, 2001, (4): 1 - 2, 9.