• 医学信息组织与利用 •

《军用医学主题词表》电子化升级研究

刘鹏年 肖 健 王天津

（军事科学院军事科学数据研究中心 北京 100039）

〔摘要〕 介绍《军用医学主题词表》基本情况、词间关系、款目格式及存在的问题，阐述运用本体工具对词表进行电子化升级的方法和过程，使其得到更便捷、直观的利用，为构建军事医学顶层本体奠定基础。

〔关键词〕 《军用医学主题词表》；电子化升级；Protege

〔中图分类号〕R－056 〔文献标识码〕A 〔DOI〕10.3969/j.issn.1673－6036.2018.03.018

Study on Electronic Upgrading of *Military Medical Thesaurus*

LIU Peng－nian, XIAO Jian, WANG Tian－jin, Information Research Center of Military Science, Academy of Military Science PLA China, Beijing100039, China

〔**Abstract**〕 The paper introduces the basic information, relation at the thesaurus level, form of entries and existing problems of the *Military Medical Thesaurus*, dilates on the method and process of the electronic upgrading of the thesaurus with ontology tools, which make it applied more conveniently and directly, to lay down foundation for the building of military medical top－level ontology.

〔**Keywords**〕 *Military Medical Thesaurus*; Electronic upgrading; Protege

1 引言

主题词表是情报检索语言的重要组成部分，是从自然语言中优选出来的规范化词典。词表作为一种术语控制工具来使用[1]，主要用于检索时的后控制和标引时自动或辅助选择索引词，目的是提高查全率和查准率，实现多语种检索和智能化概念检索。利用已有的纸质组织系统（辞典、分类法、叙词表等）快速构造本体的方法，已经成为图书情报界的共识。每部专业主题词表凝聚编表专家的大量专业知识，具有重视词汇控制，词汇标准规范，在词类控制、词形控制方面规范性强，包含英汉对照，词间关系控制良好等优点。本研究旨在通过对《军用医学主题词表》的本体化升级，将《军用医学主题词表》[2]文本格式转换为本体格式，在保留原有主题词表的结构和概念间关系的基础上，适当进行电子化扩展，使该词表可以被更便捷、直观的利用。

2 《军用医学主题词表》简介

2.1 基本情况

《军用医学主题词表》是我国军队一部大型专业主题词表，是《军用主题词表》系列的重要组成部分。该词表的编制对推动军队医药卫生事业发展和医学科学技术进步起到了重要作用。统一和规范医药卫生用语，为军队各级卫生部门、医疗、预防、教学和科研单位进行文献标引、检索和实现信息资源共享奠定基础。词表共收词 23 387 条，其中正式主题词 20 662 条，非正式主题词（入口词）

〔修回日期〕 2017－09－01

〔作者简介〕 刘鹏年，馆员，发表论文 30 余篇。

2 725条，收词以名词为主。

2.2 词间关系

2.2.1 概述 主题词表明确显示主题词间、主题词与非主题词间之间的关系以区别各词在词表中的功能与作用，方便词表使用者正确选用主题词和扩充查词，引导标引人员正确选词和用户进行扩检和缩检。常见词表有 3 种参照子系统，即用代关系（优选关系）、属分关系（属种关系）和相关关系，这 3 种词间关系在显示时相互对应[3]。《军用医学主题词表》同样遵循这一关系。

2.2.2 用代关系 是主题词和非正式主题词之间的关系，主题词用于标引和检索，非正式主题词（入口词）仅用于检索。表示用代关系的符号如下：（1）“Y”（用）：此符号后面的词为正式主题词。（2）“D”（代）：此符号后面的词为非正式主题词。

2.2.3 属分关系 是正式主题词之间的概念等级关系，具有这种关系的两个主题词，其中之一是上位词，概念较宽泛且专指度浅；另一个是下位词，概念较具体且专指度深。表示属分关系的符号如下：（1）“S”（属）：此符号后面的主题词为上位词。（2）“F”（分）：此符号后面的主题词为下位词。（3）“.”（黑点）：表示此符号后面的主题词的等级关系。收入统一词族的主题词，按其等级概念作阶梯式排列。（4）“Z”（族）：此符号后面的主题词为族首词，表示在词族中概念最大的主题词。族首词多是主要研究对象、研究方法及仪器设备等的类名称。

2.2.4 相关关系 是两个正式主题词之间的一种相互参照关系。如交叉概念、因果关系、对立统一和相互矛盾概念等主题词之间，均可建立相关关系。表示相关关系的符号：“C”（参）：此符号后面的主题词为参照词。

2.3 款目格式

一条主题词款目构成了一个概念及其各项属性的完整描述，《军用医学主题词表》中的主题词款目由以下元素构成：汉语拼音、款目词、族首词、英文译名、范畴号、含义注释、用关系词（参照项“用”）、代关系词（参照项“代”）、属关系词（参照项“属”）、分关系词（参照项“分”）、组关系词（参照项“族”）及参关系词（参照项“参”）。《军用医学主题词表》通常包括：汉语拼音、款目词、英文译名、参照项和范畴号等内容，涉及以下 4 种具体示例：（1）族首词款目格式。族首词后面带有“＊”号，参照项中可以有“分” “代” “参”，没有“属”项。（2）族中次款目格式。参照项中“分”“代”“属”“族”等内容均有。（3）无关联词款目格式。无关联词一般即无属分关系又无用代关系，参照项为“带”或“参”项。（4）非正式主题词款目格式。非正式主题词多是入口词，用来指引正式主题词，在使用上不用于标引和检索文献。参照项为“用”项。

2.4 存在的问题

现有《军用医学主题词表》是 1991 年编辑出版的，存在较多不适应现今数字化时代应用的问题。（1）人工维护成本高，需要大量的专家和人员维护。（2）非主题词（入口词）数量偏低，词表以主题词为主，入口词数量少且不适用于普通用户的使用习惯。（3）修订更新时效性差，词表建成 20 多年来，没有进行必要的修订改版，主题词表的修订工作处于停滞状态。（4）在专业使用方面，由于采用人工编制的方法，所编主题词的概念含义专指度普遍较浅，在使用上经常需要借助上位概念或多次组配，增加主题词的使用频率，也降低用户使用该词表的积极性。（5）在词表的通用共享方面，出版的词表是纸质版，使用范围封闭，其组织体系与其他主题词表（叙词表）组织体系的通用性很差，非图书情报专业人员对该词表基本不知晓等。以上这些问题都严重阻碍《军用医学主题词表》发挥其应有的作用。

3 词表电子化

3.1 概述

传统的手工编制和纸质服务的方式显然不能满足大数据时代用户对主题词表的需求，《军用医学

主题词表》在电子化、网络化发展方面，与国际国内的先进水平有很大的差距。本研究通过对《军用医学主题词表》主表中叙词款目的电子化升级，在为用户提供数字化交互式支持的基础上，解决上述存在的问题，为构建军事医学顶层本体夯实基础。

3.2 语义置标语言

本体（Ontology）是针对某一主题规范的说明。网络本体语言（Web Ontology Language，OWL）是一种用于在语义网上发布和共享本体的语义置标语言[4]。本研究选用 OWL 来表示词表，可以利用 OWL 丰富的描述机制和良好的推理能力来实现主题词表本体的一致性检测、语义关系扩展和与其他本体化的中文叙词表（主题词表）实现映射与集成[5]。

3.3 本体转换规则

本研究参照中文叙词表本体共建共享系统（OTCSS）[6]，设计制定词表转换为 OWL 本体的转换、扩展规则，见表 1。美国斯坦福大学开发的 Protégé 软件[7]国内外应用广泛，开发界面友好，插件扩展丰富，本研究采用 Protege4.3 作为本体编辑工具，完成相关主题词表到领域本体的转换。

表 1 OWL 本体转换规则

属性	值域	属性特征	标签	叙词表对应标识
HasNTerm	NTerm	–	入口词	代 D
Broader	Concept	具有传递性，与 Narrower 互逆	上位词	属 S
Narrower	Concept	具有传递性，与 Broader 互逆	下位词	分 F
TopConcept	Concept	–	族首词	族 Z
Related	Concept	–	参见	参 C
PinYin	&rdfs	只能出现一次	汉语拼音	–
EngCounterpart	&rdfs	–	英译名	–
ScopeNote	&rdfs	–	范畴注释	注

3.4 转化结果展示

转换过程中为更高效地构建领域本体，将《军用医学主题词表》进行 XML 格式的转化，根据定义的转化规则，设计 XML 文本格式。利用 Protégé 的导入功能将 XML 格式文本导入并生成 OWL 格式文件。词表可视化效果，见图 1。

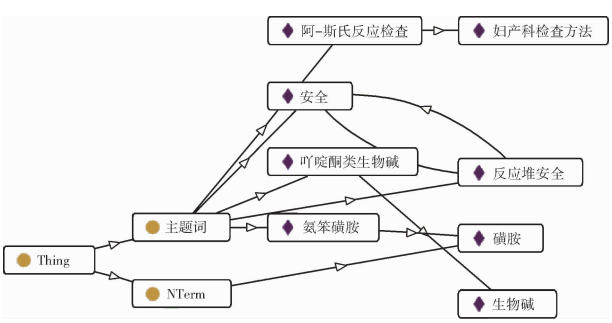

图 1 词表可视化效果

4 讨论

4.1 效果

李景[8]等从术语规范性、知识组织结构、开放性、语义关系丰富程度、知识库属性和修订便利程度等 6 个方面，对主题词表（叙词表）和本体的区别进行全面精辟的论述。通过将纸质版《军用医学主题词表》电子化升级，达到以下效果：（1）解决词表的形式化表示和细粒度关系扩展问题，电子化后的词表同时具备传统中文主题词表（叙词表）和本体的特征。（2）解决纸质版主题词表不易共享使用的问题，为实现主题词的网络化、数字化奠定良好的基础。（3）解决纸质版主题词表无法及时更新和维护的问题，新出现的词汇与关系可以方便及时地更新到电子化后的词表中。

4.2 不足

本研究仅将原有纸质版主题词表进行简单的电子化改造，还有很多不足之处：（1）为尽量保留原《军用医学主题词表》在词汇控制方面的成果，本研究在电子化转换的过程中，保留原有主题词表的结构和组成部分，在新词的吸收和新关系的创建上创新度不够。（2）在语义关系的揭示上，没有进一步细分更具体的关系，需要在今后的本体化工作中，编制比较详细的相关关系，以弥补语义网中横向关系的不足。

5 结语

本体的开发和完善是一个迭代螺旋上升的过程，本研究在《军用医学主题词表》电子化的过程中，保留原有概念体系和知识结构，建成的电子化主题词表还需在后续的升级中补充、细化和完善，最终形成军事医学顶层本体。

参考文献

1 中华人民共和国国家标准. GB/T 3860 - 1995，文献叙词标引规则［M］. 北京：中国标准出版社，1995.

2 军用医学主题词表［M］. 北京：人民军医出版社，1993.

3 国防科学技术叙词表［M］. 北京：军事科学出版社，1991.

4 OWL 2 Web Ontology Language Document Overview (Second Edition). W3C Recommendation［EB/OL］.［2017 - 12 - 11］. https://www.w3.org/TR/owl2 - overview/.

5 鲜国建，赵瑞雪，朱亮，等. 农业科学叙词表的 SKOS 转化及其应用研究［J］. 现代图书情报技术，2012，32（10）：16 - 20.

6 曾新红等. 中文知识组织系统 - 语义描述、共建及共享服务［M］. 北京：化学工业出版社，2016：26 - 31.

7 Protégé［EB/OL］.［2016 - 07 - 06］. http://protege.stanford.edu/products.php#.

8 李景，钱平. 叙词表与本体的区别与联系［J］. 中国图书馆学报，2004，30（1）：38 - 41.

9 肖健，刘伟，刘鹏年，等. 军事医学本体概念获取方法研究［J］. 中华医学图书情报杂志，2016，25（5）：21 - 25.

10 黄华军，曾新红，林伟明，等. 中文知识组织系统形式化语义描述标准体系研究（二）——分类法共享服务系统 CLSS 研究与实现［J］. 中国图书馆学报，2015，41（2）：17 - 28.

11 薛建武，赵娜，王东娜. 面向本体构建的叙词表词间关系细化和应用研究［J］. 现代图书情报技术，2013，39（3）：14 - 20，8.

12 邱琳，郑怀国，李光达，等. 基于本体构建的农业网络叙词表的编制［J］. 安徽农业科学，2011，（3）：1847 - 1849.

13 刘锦秀，丁如龄.《军用医学主题词表》评价及改进意见［J］. 情报学报，1996，15（4）：278 - 285.