

生物、医学、农业领域《汉表》顶层本体框架构建研究

中国科学技术信息研究所

2020. 10. 26

目 录

1. 引言	1
1.1 构建《汉表》顶层本体的必要性	1
1.2 构建《汉表》顶层本体的指导原则	2
2. 生物、医学、农业领域《汉表》体系结构	3
2.1 设计原则	3
2.2 体系结构	3
2.3 顶层本体的构成	4
2.4 顶层本体的作用机制	6
2.5 顶层本体的构建参考	8
2.6 顶层本体示例—UMLS 语义网	8
3. 顶层本体语义类型的构建	10
3.1 语义类型的设立准则	10
3.2 语义类型的“subclassof”关系	14
3.3 语义类型的筛选与调整	14
3.3.1 将 UMLS 语义类型进行简化	14
3.3.2 其他词系统顶层概念与 UMLS 语义类型的映射	16
3.3.3 语义类型的调整	25
3.4 语义类型框架的检验	28
3.5 语义类型框架的不足	29
4. 顶层本体语义关系的构建	30
4.1 语义关系的筛选	30
4.1.1 UMLS 语义关系统计分析与筛选	30
4.1.2 其他词系统语义关系与 UMLS 语义关系的映射	34
4.2 语义关系的调整	42
5. 顶层本体副主题词的构建	45
参考文献	48
附表	49

插图清单

图 1 生物、医学、农业领域《汉表》体系结构.....	4
图 2 顶层本体的构成及作用机制.....	4
图 3 MeSH 中“肝脏”可组配的副主题词.....	7
图 4 部分语义类型树形结构示例.....	9
图 5 生物分类演变对比图.....	11
图 6 UMLS 前三级语义类型结构框架.....	15
图 7 调整后的语义类型框架图.....	27
图 8 UMLS 语义网中相关关系连接的语义类型组数.....	32
图 9 副主题词表层级结构图（其中黑色字体表示来源于 MeSH，蓝色字体表示来源于中国中医药学主题词表，红色字体表示来源于中国药学主题词表）.....	47

附表清单

表 1 UMLS 中按概念数排列的语义类型（前 30 种）.....	15
表 2 AGROVOC 顶层概念与 UMLS 语义类型映射结果.....	17
表 3 CAB 顶层概念与 UMLS 语义类型映射结果.....	18
表 4 MeSH 顶层概念与 UMLS 语义类型映射结果.....	19
表 5 SNOMED CT 顶层概念与 UMLS 语义类型映射结果.....	22
表 6 中医药学主题词表顶层概念 UMLS 语义类型进行映射结果.....	24
表 7 5 个词系统顶层概念与 UMLS 语义类型的映射结果统计.....	25
表 8 随机抽取的 100 个生物、医学、农业词汇的语义类型分配结果.....	28
表 9 UMLS 的 53 种相关关系.....	30
表 10 UMLS 语义网中连接语义类型较多的 35 种语义关系.....	33
表 11 UMLS 英文来源词汇表按使用频次排列的 32 种关系类型.....	34
表 12 UMLS 来源词汇表的主要相关关系与 UMLS 语义关系的映射.....	35
表 13 AGROVOC 主要相关关系与 UMLS 语义关系的映射.....	36
表 14 《国家标准 GB/T 13190.1-2015》中代表性相关关系与 UMLS 语义关系的映射.....	37
表 15 SNOMED CT 中术语 Closed fracture of radius AND ulna (disorder) 的相关关系描述.....	38
表 16 SNOMED CT 中主要相关关系的使用频次.....	39
表 17 SNOMED CT 中主要相关关系与 UMLS 语义关系的映射.....	40
表 18 语义关系结构框架表.....	42