

SNOMED CT 概念关系表达与语义检索^{*}

夏光辉 李军莲 李晓瑛 李丹亚

(中国医学科学院医学信息研究所 北京 100020)

〔摘要〕 详细阐述 SNOMED CT 基于概念的组织框架、表示模式、关系模型、表达规则等, 初步探索 SNOMED CT 在医疗数据表达、语义检索方面的应用, 为研究制定我国临床诊疗术语标准提供参考借鉴, 进一步推动临床医疗数据的处理、挖掘与分析研究。

〔关键词〕 SNOMED CT; 概念关系; 语义关系; 概念表达; 语义检索

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2017.03.011

Concept Relation Expression and Semantic Retrieval of SNOMED CT XIA Guang-hui, LI Jun-lian, LI Xiao-ying, LI Dan-ya, Institute of Medical Information, Chinese Academy of Medical Sciences, Beijing 100020, China

〔Abstract〕 The paper describes the organization framework, representation pattern, relation model and expression rule of SNOMED CT based on concept, and then explores the application of SNOMED CT in the expression of medical data and semantic retrieval. Thus, it can provide reference for the research and development of the terminology standard of clinical diagnosis and treatment, and promote the study of processing, mining and analysis of clinical medical data in China.

〔Keywords〕 SNOMED CT; Concept relation; Semantic relation; Concept expression; Semantic retrieval

1 引言

系统化临床医学术语集 (The Systematized Nomenclature of Human and Veterinary Medicine, Clinical Terms, SNOMED CT) 是由美国病理学会编著出版的当今世界上最庞大的医学术语集, 是在美国病理学会编制的 SNOMED RT 与英国国家卫生服务部 (National Health Service, NHS) 编制的临床术语集

第 3 版 (Clinical Terms Version 3, CTV3) 相互融合的基础上, 经过扩充和结构重组而形成的^[1], 从最初 SNOP 采用的 4 轴分类体系扩充到目前 SNOMED CT 的 19 轴分类体系, 其用途也从最初满足以病理学术语为主体内容的分类和检索扩展为能够满足信息化时代临床信息系统的检索、数据汇聚、分析和交互共享。

2007 年 4 月, 美国病理学会将 SNOMED CT 的维护、发布、知识产权转让给国际医疗术语标准发展组织 (International Health Terminology Standards Development Organization, IHTSDO)。IHTSDO 是一个总部设在丹麦的非营利性组织, 其建立时成员包括澳大利亚、加拿大、丹麦、立陶宛、荷兰、瑞典、新西兰、美国 and 英国 9 个国家, 目前其成员国已经包括 28 个国家和地区^[2]。作为拥有术语使用许可权的美国病理学会, 在与 IHTSDO 的合同约定

〔修回日期〕 2016-11-11

〔作者简介〕 夏光辉, 助理研究员, 发表论文 20 余篇。

〔基金项目〕 中国医学科学院中央级公益性科研院所基本科研业务费项目“基于中文病例报告的命名实体识别与应用研究”(项目编号: 2016ZX330018)。

下继续对 SNOMED CT 进行研究与开发, 继续提供与 SNOMED CT 相关的产品及服务。目前 IHTSDO 公布使用 SNOMED CT 有两种途径: 一是成员国免费使用; 二是非成员国付费使用^[3]。由于我国暂时还没有加入 IHTSDO, 因此 SNOMED CT 在国内临床诊疗数据挖掘方面的应用还处于探索阶段, 但 SNOMED CT 基于概念的组织方式、概念间语义关系以及概念表达形式对于今后我国开展临床诊疗术语标准化研究, 具有较大借鉴意义和实践价值。

2 概念表示及关系模型

2.1 概述

概念、关系和描述是 SNOMED CT 最基本的组成单元。其中, 概念表不仅收录了具有明确临床意义的术语的规范表达形式, 而且还收录了表达概念之间关系的规范关系名称; 描述表收集了各个概念的不同表达形式, 对各种表达形式的用途进行了划分; 关系表揭示了各个概念之间的语义关系。SNOMED CT 采取数字标识符来唯一地表达一个概念、描述和关系, SNOMED CT 概念、描述和关系间的逻辑模型结构, 见图 1。

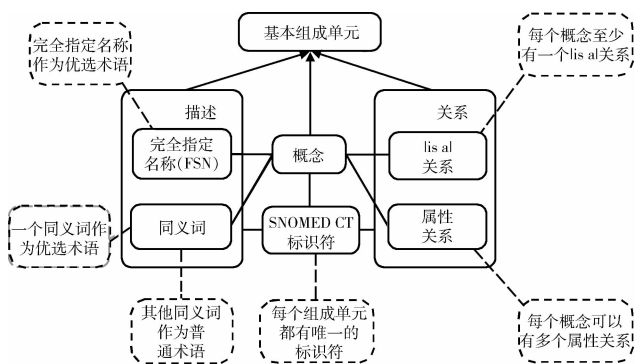


图 1 SNOMED CT 逻辑模型结构

2.2 概念表示模式

在 SNOMED CT 中, 概念是一个具有明确临床意义的标识, 由一个唯一的数字标识符 (ConceptID) 表示。SNOMED CT 的每个概念包含一组术语集, 同时通过基于“语义”的概念关系来逻辑化定义概念, 用数字编码固化每一个概念拥有的多个同义术语及具有的纵、横向关系, 以利于计算机理

解、处理和查询。SNOMED CT 依据现代西医学本体论疾病观将所有概念分为 19 个顶级类, 全面覆盖了临床科研工作所需的概念种类, 各类之间具有不同的逻辑关系, 通过“IS_A”关系建立类之间的等级体系, 其中“临床发现”与“操作”是核心类, 其他类为“支持性类”^[4]。SNOMED CT 中的概念通过完全指定名称来准确表示, 完全指定名称由概念名称和概念语义类型组成, 每一个概念的结尾都会在括号内标识概念的语义类别, 以此来区分概念在不同语义环境下表示的不同含义, 每个完全指定名称以“概念名称 (语义类型)”的形式构成, 如 Tuberculosis (disorder) — 肺结核 (紊乱)。SNOMED CT 的概念共赋予了 41 种语义类型, 每个概念通过语义类型可以明确区分相同术语形式表达的不同含义。

2.3 概念关系模型

SNOMED CT 以概念为核心, 通过属性 (也称关系) 表达来源概念和目标概念之间的关联关系, 每一个属性其本身也是 SNOMED CT 的一个概念。SNOMED CT 的概念关系划分为定义 (Defining)、修饰限定 (Qualifying)、历史 (Historical) 和附加 (Additional) 4 种类型。每一种概念关系都通过特定的属性来表达, 形成“概念 1 + 属性 + 概念 2”的三元组结构。三元组结构不仅可以用来表达一个明确的临床事物, 而且能够揭示出两个概念间语义关系的具体含义, 三元组结构中的概念 2 实质是对概念 1 的属性进行赋值。定义类属性是实现概念逻辑化定义的主要手段, SNOMED CT 20160131 国际版共包含了 66 种属性, 其中 58 种为定义类属性^[5]。SNOMED CT 的属性用于表示一个概念的特性, 其通常可以应用于一个概念类的层级类目, 有少数可应用于多个概念类的层级类目。一个属性可以应用的层级类目被称为属性的“域” (Domain)。每个属性可以被赋予一组有限的值, 这个值域集合被称为属性的“值域” (Range)。例如, | 发现部位 | 的“域”是 | 临床发现 (发现) | 及其子类; | 发现部位 | 的“值域”是 | 解剖结构 (身体结构) | 或 | 获得身体结构 (身体结构) | 及其子类。以属性“发现部位”为例, 其对应的“域”为肺炎, 对应“值域”

为肺组织，即可构建概念关系三元组“|肺炎| - |发现部位| - |肺组织|”，见图 2。

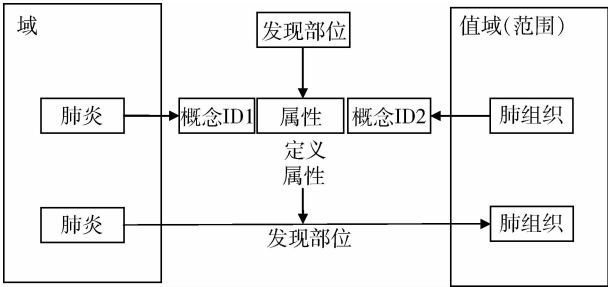


图 2 |肺炎| - |发现部位| - |肺组织| 的概念关系表示

SNOMED CT 中共有 59 个属性明确规定了应用的“域”，包括 57 个定义类属性和 2 个修饰限定类属性，其中“临床所见”16 个、“操作”23 个、“评估操作”6 个、“具有明确上下文语境”6 个、“事件”5 个、“标本”5 个、“药品/生物制品”2 个、“身体结构”1 个、“物品”1 个。同一个属性也可以用于不同的“域”，如应用于“事件”概念类的 5 个属性同时可用于“临床所见”；属性“|has active ingredient|”（有活性成分）可同时用于“物品”和“药品/生物制品”概念类。“临床所见”域的部分属性及其对应值域，见表 1。

表 1 “临床所见”域的部分属性及其对应值域

域	属性	值域
(临床所见) Clinical Finding	发现部位 (Finding Site)	(解剖结构) Anatomical structure
		(获得性体结构) Acquired body structure
	相关形态 (Associatd Morphology)	(形态结构异常) Morphologically abnormal structure
		与...相关 (Associated With)
		(临床所见) Clinical Finding
		(操作) Procedure
		(事件) Event
		(有机体) Organism
		(物质) Substance
		(物品) Physical object

续表 1

	(物理力) Physical force
	(药品/生物制品) Pharmaceutical/biologic product
致病因子 (Causative Agent)	(有机体) Organism
	(物质) Substance
	(物品) Physical object

3 概念表达

3.1 概述

SNOMED CT 以先组概念和后组概念表达方式来保证临床事物能够被正确地描述。先组概念是指 SNOMED CT 概念表中预先定义的概念，每一个先组概念不仅具有唯一的概念标识符和不同表达形式，同时通过定义类关系实现与相关概念的关联，以明确表达概念的逻辑定义。后组概念是指通过两个以上先组概念的组合作所表达的一个概念，通过后组概念的表达式，可以利用上位概念来表达其下位概念的具体含义。后组概念表达方式主要有两方面的意义：一是可以有效地减少先组概念的数量，通过后组概念的方式来表达相同含义的概念。二是可以利用先组概念来表达颗粒度更细、含义更具体的新概念。

3.2 后组概念表达规则

IHTSDO 为 SNOMED CT 制定了清晰的概念组配语法规则，以支持后组概念的表达式既能够便于计算机自动处理，同时也符合人类语言的表达习惯。在后组概念表达过程中，必须要遵循先组概念的概念关系模型，属性只能应用于相应的“域”和“值域”。具体表达规则如下：(1) 概念标识符是概念表达规则的基本组成单元。(2) 概念也可用概念标识符联合概念名称的形式表达，概念名称首尾用“|”进行分隔。(3) 主体概念与关系概念之间用“:”分隔。(4) 关系概念与客体概念之间用“=”分隔。(5) 主体概念可组配多个关系概念，关系概

念与其对应的客体概念形成一个组合,各组合之间用“,”分隔。(6)关系概念划分为不同类型时,相同类型的关系概念以及其对应的客体概念通过大括号“{}”分隔。(7)客体概念本身可通过后组概念形式来表达,用小括号“()”分隔客体概念来表达嵌套结构。

3.3 表达相同含义的先组概念

当 SNOMED CT 先组概念具有完整定义时,概念的内涵可以通过定义属性划分为多个方面进行详细阐述,同时此先组概念也可以通过多个方面的组合来表达。例如“腹腔镜阑尾紧急切除术”(174041007 | laparoscopic emergency appendectomy |)在 SNOMED CT 中为先组概念,其包含阑尾切除、使用腹腔镜、紧急处理等方面,可通过定义属性描述其具体的含义:

174041007 | laparoscopic emergency appendectomy | : 116680003 | is a | = 80146002 | appendectomy | (腹腔镜阑尾紧急切除术是一种阑尾切除术)

174041007 | laparoscopic emergency appendectomy | : 260870009 | priority | =

25876001 | emergency | (腹腔镜阑尾紧急切除术适用于紧急情况)

174041007 | laparoscopic emergency appendectomy | : 425391005 | using access device | = 86174004 | laparoscope | (腹腔镜阑尾紧急切除术使用腹腔镜)

同时也可以通过其上位概念“阑尾切除术”(80146002 | appendectomy |)的后组概念表达形式来等价表达:

80146002 | appendectomy | : 260870009 | priority | = 25876001 | emergency | , 425391005 | using access device | = 86174004 | laparoscope | (紧急情况下进行的阑尾切除术 & 使用腹腔镜进行的阑尾切除术)

3.4 表达新概念

后组概念表达最主要的用途是表示 SNOMED CT 先组概念中不存在的概念,通过后组概念表达方式,就可以利用先组概念精确地表达含义更为具体的概念。例如 SNOMED CT 没有先组概念“腹部腹腔镜拆除装置”(laparoscopic removal of device from

abdomen),可采用后组概念表达方式,通过先组概念“腹部拆除装置”(68526006 | removal of device from abdomen |)进行等价表达:

68526006 | removal of device from abdomen | : 425391005 | using access device | = 6174004 | laparoscope | (腹部拆除装置时使用腹腔镜)

后组概念表达依据关系表中已经存在的定义类属性,当 SNOMED CT 的概念关系表中使用一个上位概念作为定义属性的“值域”时,可以选择该概念的下位概念作为定义属性的“值域”,就可以表达一个更加具体的概念。例如 SNOMED CT 可以通过后组概念形式表达概念“全髋关节假体置换术”:

52734007 | Total replacement of hip | : 363699004 | direct device | = 304120007 | total hip replacement prosthesis | (使用全髋关节置换假体进行全髋关节置换术)

此时也可以选择先组概念“全髋关节置换假体”(304120007 | total hip replacement prosthesis |)的下位概念“Exeter 全髋关节假体”(390783005 | Exeter total hip prosthesis |)作为定义属性“直接设备”(direct device)的“值域”,以表达“Exeter 全髋关节假体置换术”这个更具体的概念:

52734007 | Total replacement of hip | : 363699004 | direct device | = 390783005 | Exeter total hip prosthesis | (使用 Exeter 全髋关节假体进行全髋关节置换术)

3.5 表达概念逻辑定义

SNOMED CT “定义”类关系可用于实现对概念进行“逻辑化”定义。SNOMED CT 的“定义”类属性由两大类组成。第 1 类是“Is a”关系,又称“上下位关系”或“父子关系”。“Is a”关系是 SNOMED CT 建立概念分类层级体系的基础,众多语义上具有包含与被包含的概念基于“Is a”关系纵向聚合,形成了 SNOMED CT 中独特的顶级概念分类层级体系。SNOMED CT 收录的全部 40 万条概念中除位于最顶级的根概念 | SNOMED CT Concept | 外,每个概念均至少与其一个上位概念通过“Is a”关系相联系,从而建立了 SNOMED CT 概念之间的层级结构。第 2 类是“概念模型属性”关系,“概念模型属性”关系与“Is a”关系相辅相成,从横向来连接两个概念并确定概念间的语义关系^[6]。

SNOMED CT 中的任何一个概念都可通过与其他概念间的关系来实现自己的逻辑化定义。以“肾疾病”为例，其逻辑化的定义文本可描述为 | Kidney disease | | is a | | Kidney finding | and | is a | | Disorder of abdomen | and | is a | | Disorder of body cavity | and | is a | | Disorder of kidney and/or ureter | and has | Finding site | | Kidney structure | （肾疾病是一种肾脏临床表现、腹部疾病、体腔疾病、肾和（/或）输尿管疾病，其发生部位为肾脏），见图 3。

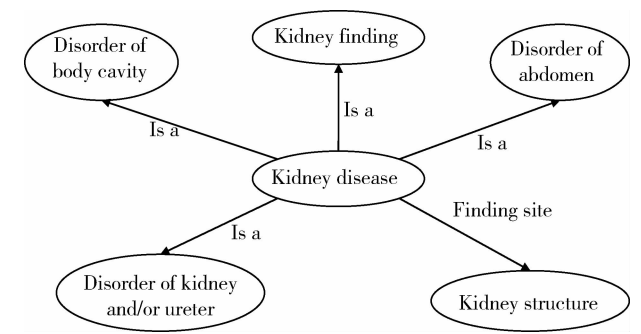


图3 概念“肾疾病”的逻辑定义

4 语义检索

利用 SNOMED CT 对医疗数据资源进行组织揭示，基于 SNOMED CT 的先组概念和后组概念表达形式，能够较好地支持基于概念的语义检索。基于 SNOMED CT 进行语义检索，不仅能够支持概念内涵的语义检索，同时也能够较好支持智能问答方式的语义检索。如查询“腹腔镜手术”的数据时，不仅可以直接查询先组概念“腹腔镜手术”，同时也可以通过主体概念“操作”、关系概念“使用设备”、客体概念“腹腔镜”的三元组组配表达方式来检索“腹腔镜手术”的相关数据。同样，如以提问的方式查询“使用腹腔镜这种设备进行手术操作”的数据时，也可以通过 SNOMED CT 的后组概念表达方式等价表达先组概念，直接以概念“腹腔镜手术”检索相关数据。如需查询病理形态为良性肿瘤的相关疾病数据时，只需要将查询条件转换为查找概念，即哪些概念的相关形态为良性肿瘤即可，见图 4。

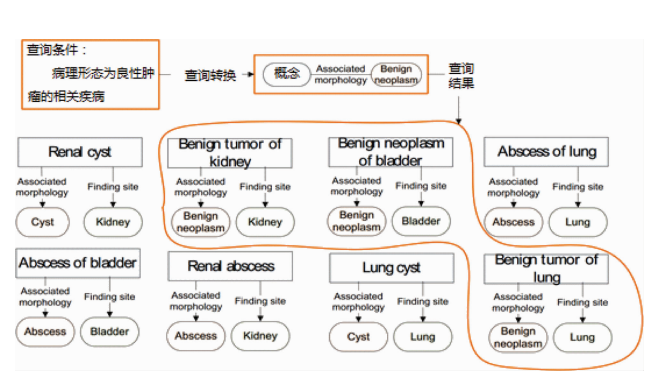


图4 查询病理形态为良性肿瘤的相关疾病

5 结语

国际医疗术语标准开发组织颁布的技术实施指南定义了 SNOMED CT 3 种类型的应用，即临床记录、知识表示、汇聚和分析^[7]。临床记录涉及病人数据和医疗服务的处理，例如在临床信息系统中，基于 SNOMED CT 术语进行记录、存储、检索和交互。知识表示是指表达临床知识，如基于 SNOMED CT 的临床指南和护理路径。汇聚和分析是指从临床信息系统中检索数据，以便于进行二次分析。目前，美国、英国、加拿大、新西兰和澳大利亚等国家已经指定 SNOMED CT 作为临床信息系统的临床术语参考标准，已经有超过 50 个国家对 SNOMED CT 开展了不同程度的应用，不仅包括对 SNOMED CT 内容的增补，如创建 SNOMED CT 的子集、扩展表、交叉映射表等；还应用于临床诊疗数据的存储、录入、检索等，以支持对诊疗数据的挖掘与分析^[8]。我国尚未成为 IHTSDO 的成员，因此本文通过对 SNOMED CT 的概念关系表达与语义检索进行研究，可为我国制定临床诊疗术语标准提供参考借鉴，进一步推动临床医疗数据的处理、挖掘与分析研究。

参考文献

- 1 IHTSDO. History Of SNOMED CT [EB/OL]. [2016 - 09 - 20]. <http://www.ihtsdo.org/snomed-ct/what-is-snomed-ct/history-of-snomed-ct>.
- 2 IHTSDO. Members [EB/OL]. [2016 - 03 - 20]. <http://www.ihtsdo.org/members>.
- 3 IHTSDO. SNOMED CT and Licensing [EB/OL]. [2016 - 09 - 20]. <http://www.ihtsdo.org/snomed-ct/get-snomed-ct>.

领域高产论文作者合著网络的密度、平均路径、凝聚子群、核心-边缘机构以及中心性,得出以下结论:整体网络分析结果表明医学伦理学领域高产作者合著网络密度较低,子群数量多但不紧密,将不同合著子群连接起来的作者也比较少,大部分高产作者不能进入网络核心位置,合著圈的范围也不广泛。网络个体分析结果说明,点度中心度大于 10 的作者仅占 13.53%,大多数作者其中间中心度值为 0,这与医学伦理学领域作者合著交流规模不广泛有关。依据本研究的阶段性结果,国外医学伦理学领域高产作者应该有意识地扩大合作交流圈,改变交流不频繁、合作范围小的状况,促进医学伦理学领域的知识共享与发展。

参考文献

- 1 许丹,侯跃芳. 国外医学信息学研究领域可视化引文分析 [J]. 医学信息学杂志, 2011, 32 (10): 12-19.
- 2 范建华,梁敏,陈明明,等. 我国缺血缺氧性脑损伤幼鼠研究文献作者合作度分析 [J]. 中华医学图书情报杂

志, 2016, 25 (5): 58-62.

- 3 丁东学. 文献计量学基础 [M]. 北京: 北京大学出版社, 1992: 204-232.
- 4 钟文娟. 基于普赖斯定律与综合指数法的核心作者测评: 以《图书馆建设》为例 [J]. 科技管理研究, 2012, 32 (2): 57-60.
- 5 邱均平,周春雷. 发文量和 h 指数结合的高影响力作者评选方法研究: 以图书情报学为例的实证分析 [J]. 图书馆论坛, 2008, (6): 44-49.
- 6 曹霞,崔雷. 基于 SNA 的国外医学信息学领域合著网络研究 [J]. 现代情报, 2016, 36 (3): 129-134.
- 7 陈少龙. 基于社会网络分析方法的合著网络在上海大学管理学院的应用 [J]. 现代情报, 2012, 32 (4): 165-168, 177.
- 8 Freeman L C. Centrality in Social Networks Conceptual Clarification [J]. Social Network, 1979, 1 (3): 215-239.
- 9 朱庆华,李亮. 社会网络分析法及其在情报学中的应用 [J]. 情报理论与实践, 2008, 31 (2): 179-183, 174.
- 10 尹莉. TopRank (k) 算法与 PageRank 算法的比较研究 [J]. 情报探索, 2014, 11: 13-15, 19.

(上接第 53 页)

- 4 IHTSDO. SNOMED CT Starter Guide [EB/OL]. [2016-09-20]. http://doc.ihtsdo.org/download/doc_StarterGuide_Current-en-GB_INT_20141202.pdf.
- 5 NIH. International Release of SNOMED CT [EB/OL]. [2016-09-20]. <https://www.nlm.nih.gov/healthit/snomedct/international.html>.
- 6 董燕,贾李蓉,朱彦,等. 2013 版 SNOMED CT 顶层概念调整及属性关系 [J]. 中国医学创新, 2014, 11 (3):

106-108.

- 7 IHTSDO. SNOMED CT? Technical Implementation Guide January 2015 International Release (US English) [EB/OL]. [2016-09-20]. http://doc.ihtsdo.org/download/doc_TechnicalImplementationGuide_Current-en-GB_INT_20150131.pdf.
- 8 Lee D, Cornet R, Lau F, et al. A Survey of SNOMED CT Implementations [J]. Journal of Biomedical Informatics, 2012, 46 (2013): 87-96.

敬告作者

《医学信息学杂志》网站现已开通,投稿作者请登录期刊网站: <http://www.yxxxx.ac.cn>, 在线注册并投稿。

《医学信息学杂志》编辑部