

RxNorm多词表语义互操作机制研究*

□ 李芳 陈颖 侯丽 冀玉静 / 中国医学科学院医学信息研究所 北京 100020

摘要: 信息系统的语义互操作是网络环境下知识组织系统建设与应用中面临的重要课题。在药学领域, RxNorm是具有代表性的语义互操作支持工具。文章通过系统分析RxNorm知识组织方式, 从概念表示、属性继承与规范化、多层次术语类型和结构化语义网以及数据存储方式等多方面来解析RxNorm多词表语义互操作的实现机制, 列举典型应用实例, 并提出对其成功经验的一些思考。

关键词: RxNorm, 语义互操作, 药学词表, RxNorm唯一概念标识符

DOI: 10.3772/j.issn.1673—2286.2012.10.012

1 引言

药品信息系统由于采用不同命名法, 不同机构之间甚至同一机构内部进行数据交换面临着很大障碍。实现系统之间互操作, 提高信息共享效率、保障患者用药安全, 亟需一个规范化的命名表作支撑^[1]。由美国国立医学图书馆(NLM)编制的RxNorm正是一个覆盖美国全境的临床药品规范化命名表^[2]。它深度整合了11个药房系统常用的大型词表, 通过规范化的名称和编码作枢纽, 有效连接了不同来源表在一定抽象水平上的同义概念。在借鉴UMLS组织方式的基础上, RxNorm采用了有自身特点表示模式和构建机制。本文通过系统分析RxNorm知识组织方式来阐述其实现多词表语

义互操作的机制, 以期为我国知识组织体系建设、卫生信息化发展提供有益借鉴。

2 RxNorm基本信息

为满足药品信息系统对语义互操作支持工具的需求, NLM自2002年在UMLS项目中开展RxNorm建设调研, 在验证其运行成本和操作结构可行性后, NLM开发了独立的RxNorm编辑和加工数据库。2004年11月RxNorm首次发布, 并确立了每月发布的日程安排。2008年10月, RxNorm在每月发布的基础上新增每周发布。经过数年的发展, RxNorm整合的词表从5个发展到11个, 收词量超过46万^[3], 已成为美国联邦政府进行临床电子数据交换指定采用的标准。

RxNorm是临床药品的规范化命名表, 同时是支持药品术语表和药学信息系统之间语义互操作的工具^[2]。它收录了美国境内可获得的处方药和相当数量的非处方药, 主要包含临床药品和药品包(后者指包含多个药品的集合, 或按一定顺序投递的药品); 未收录的是放射性药物、造影剂、食物、食品补充剂和医疗用具(如绷带和拐杖)等^[4]。

3 RxNorm多词表语义互操作实现机制

语义互操作(Semantic Interoperability)是指系统间共享的数据能够在充分定义的领域概念水平上得到理解的能力^[8]。在卫生保健领域, 由于各信息系统采用不同名称和代码形式, 信息交互与

* 本文系国家“十二五”科技支撑计划课题“面向外文科技文献的超级科技词表和本体建设”(编号: 2011BAH10B01)子任务的研究成果之一。

共享面临难以逾越的障碍。单一、标准、多目的药品术语表是解决语义互操作的关键基础^[3]。

RxNorm由UMLS的核心团队构建,其宏观思路借鉴了UMLS基本原理^[5],包括以多个来源表为数据源,通过概念组织数据,为概念分配CUI、添加属性和构建概念间关系,形成语义网,并以RRF格式(Rich Release Format)文档保存数据,达到对多个异构词表的有效整合。以建设一个规范化命名表来促进药品信息系统互操作为目的,RxNorm知识组织方式又有自身特点。它通过机器辅助的人工干预方式进行同义概念归并、歧义概念处理,生成规范化概念名称和概念、唯一标识符(RxNorm Concept Unique Identifier, RxCUI),继承与规范化来源表重要编码,构建多层次的术语类型和结构化语义网,支持含不同药学词表的计算机系统之间进行高效和无歧义的数据交互(如图1所示)。

3.1 规范化概念名称和 RxCUI是实现多系统互操作的枢纽

为支持临床药品系统的互操作,RxNorm选用临床处方形式作为数据表示模式。为达到语义层面的交互,RxNorm通过概念组织数据,界定各来源表的同义概念,为其提供语义规范化形式(Semantic Normal Form, SNF)的名称和概念唯一标识符 RxCUI,使得一定抽象水平上的同义概念都能连接到同一个标准化名称和代码,是RxNorm实现信息系统语义互操作的基本机制。

(1) 概念表示模式

在咨询HL7和退伍军人健康管理局(VA)技术委员会基础上,NLM选用临床医师和药师所接受的临床处方形式作为概念表示模式^[1],即以“活性成分+剂量+剂型[商品名]”来表示临床药品,如“氟西汀 4MG/ML口服液[百忧解]”。当

其中任何一个元素发生变化时,就会有一个新的RxNorm概念产生。RxNorm命名表包含临床有意义化学成分每种剂型与剂量的可获得组合^[2]。RxNorm为每个概念分配唯一标识符RxCUI,它不会被删除或重用,并在RxNorm各版本中得到延续。

(2) 同义概念归并与语义规范化名称生成

RxNorm遴选美国境内药房管理和药品相互作用系统最常用的11个词表作为药品名称来源(如表1所示)。

作为临床药品标准命名表,RxNorm识别各来源表同义词并为其生成规范化形式。由于概念表示模式的特殊性和实际数据的复杂性,RxNorm对同义概念的归并主要采用机器辅助的人工干预方式。例如,对来自不同词表的数据^[2]:

- Naproxen Tab 250 MG
- Naproxen 250mg tablet (product)
- NAPROXEN@250 mg@ORAL@TABLET
- Naproxen 250 MILLIGRAM In 1 TABLET ORAL TABLET
- NAPROXEN 250MG TAB,UD [VA Product]

尽管表示形式不同,编辑人员认为它们在一定抽象水平上有相同的意思,可作为同义概念归并在一起,生成“Naproxen 250 MG oral tablet”作为其规范名称,并分配共同的RxCUI“198013”。

经过人工干预、衍生生成的规范化药品名称集合构成RxNorm规范术语集(SAB=RxNorm)。相对于RxNorm(来源词表数据与生成数据的总和),RxNorm规范术语集仅指衍生生成的规范数据部分^[2]。

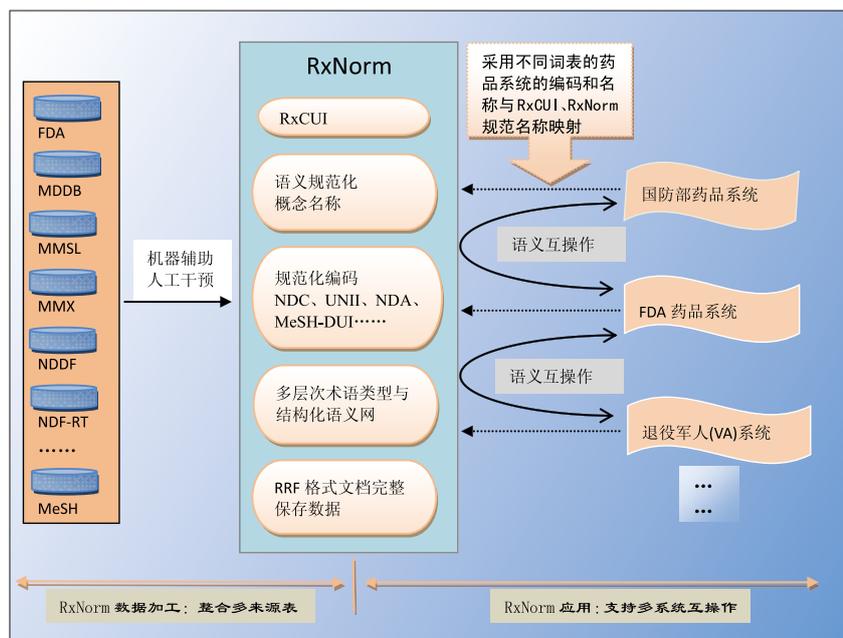


图1 RxNorm多词表语义互操作机制

表1 RxNorm来源词表

词表全称 ^[2]	缩写形式 ^[2]	从各来源表所收录词数 ^[6]
MultumMediSource Lexicon	MMSL	36,248
FDB MedKnowledge (formerly NDDF Plus)	NDDF	25,127
Micromedex RED BOOK	MMX	23,638
Veterans Health Administration National Drug File - Reference Terminology	NDFRT	16,517
FDA National Drug Code Directory	MTHFDA	16,213
Veterans Health Administration National Drug File	VANDF	16,049
Gold Standard Drug Database	GS	14,891
FDA Structured Product Labels	MTHSPL	14,710
SNOMED Clinical Terms (drug information)	SNOMEDCT	13,007
Medi-Span Master Drug Data Base	MDDB	8,241
Medical Subject Headings (MeSH)	MSH	6,974

(3) 歧义概念处理^[7]

RxNorm数据来源于多个词表,各词表更新与调整周期不同,同形或相似的字符串可能有不同含义,需要将它们作为歧义概念进行处理。

形成歧义的主要原因包括:

(1) 歧义是不能分辨字符串准确含义的结果。例如一个原词的剂型部分是“CAP/TAB”,则其剂型可能是胶囊剂(capsule)或片剂(tablet),造成歧义。(2) 歧义是由数量的多种可能性导致。例如对

于某种品牌的吸入器,可能有多种使用次数,但原词本身未给出具体数量信息,形成歧义。

RxNorm通过复制的方法来处理歧义字符串。将原始词作为基础词(base atom),不为其分配RxNorm形式;复制词(duplicate atom)用原始词加上‘_#N’(‘N’为第1到第N个复制词)表示,生成RxNorm规范形式反映该复制词的准确含义,并为每个复制词分配RxCUI。对“歧义标记

(AMBIGUITY_FLAG)”属性,原始词和复制词的属性值分别为“Base”和“Duplicate”。除了NDC编码和歧义标识属性,复制词与基础词有同样的关系和属性。

例如,对于商品名为“Xopenex”的沙丁胺醇(Levalbuterol)吸入剂,有80和200两种作用次数,其基础词和复制词的属性和规范名称(如表2所示):

NLM常规性地认真检查RxNorm可能存在的歧义字符串,以保障RxNorm规范形式在一定抽象水平上意思是明确的,即RxNorm规范术语集不会存在歧义。

RxNorm采用人工干预的方式充分理解来源表术语的原本含义,通过同义概念的归并和歧义概念的拆分实现术语概念的准确定位,为无歧义的同义概念创造语义规范化的概念名称和RxCUI。当不同名称和编码有共同的RxCUI时,它们可以相互映射。

3.2 RxNorm丰富与规范化的属性是多系统互操作的有益支撑

属性(Attribute)是与概念或原词有关的其他信息片段。RxNorm充分继承来源表属性,同时添加新的相关信息。其中,对来源表重要编码(CODE)属性的继承和规范化是支持语义互操作的又一重要机制。通过CODE与RxCUI的映射,或规范化CODE之间的映射,实现了采用不同命名表或不同编码体系的临床信息系统之间的有效交互。

(1) 继承与规范来源表重要编码^[2]

RxNorm规范术语集继承了

表2 “沙丁胺醇”歧义字符串的处理

表示形式	歧义属性	RxCUI	RxNorm规范名称
Levalbuterol 0.045 MG/ACTUAT Metered Dose Inhaler [Xopenex]	Base	996225	无
Levalbuterol 0.045 MG/ACTUAT Metered Dose Inhaler[Xopenex]_#1	Duplicate	996233	80 ACTUAT Levalbuterol 0.045 MG/ACTUAT Metered Dose Inhaler[Xopenex]
Levalbuterol 0.045 MG/ACTUAT Metered Dose Inhaler[Xopenex]_#2	Duplicate	746466	200 ACTUAT Levalbuterol 0.045 MG/ACTUAT Metered Dose Inhaler[Xopenex]

各来源词表的重要编码信息,如MeSH叙词标识符DUI、FDA唯一成分标识符(UNII)、新药申请号(NDA)以及国家药品编码(NDC)等。其中,NDC由多个来源词表提供并且形式多样,RxNorm规范术语集将其转化为11位HIPAA格式,作为NDC属性进行存储。

(2) 添加属性^[2]

除了继承来源表信息特征,RxNorm规范术语集添加了自身许多属性特征,包括剂量、单一成分或多成分、是否含有精确成分、作用时间/次数等,以及过期药品名称的相关属性描述。

(3) 整合NDF-RT的药品临床属性^[3]

由退伍军人健康管理局(VA)编制的国家药品文档参考术语集(NDF-RT)提供了药品相关的临床属性,包括可能的临床用途、药理属性(作用机制或生理效果)、化学结构、使用禁忌以及药物相互作用。RxNorm自2010年6月将NDF-RT整合进入系统。

3.3 多层次术语类型和结构化语义网促进不同表现形式的药品信息交互

在实际应用中,不同人群在不同时期对药品表示形式的需求是多样的。例如,对于临床医生,在写医嘱时认为使用处方表达形式是方便的,而在药物相互作用或过敏反应检查时,仅需要药品活性成分;对药师来说,商品名以及连接商品名和通用名的关系是有用的;对于患者,通常只记得商品名或通用名,而记不住准确的剂量或剂型^[3]。RxNorm通过术语类型

(Term Types, TTYs)来组织不同层次的概念描述,以满足不同的信息需求。

(1) 术语类型^[2]

主要包括通用名(Generic Names)和商品名(Branded Names)两大类型(如表4所示),通用名形式包括IN(成分)、SCDC(成分+剂量)、SCDF(成分+剂型)、SCD(成分+剂量+剂型)以及通用名药物包,商品名形式包括BN(商品名)、SBDC{成分+剂量[商品名]}、SBDF{成分+剂型[商品名]}、SBD{成分+剂量+剂型[商品名]}以及商品名药物包,支持用户从化学成分、商品名、通用名等不同抽象水平获取药品信息。

(2) 关系与语义网^[1]

SAB=RxNorm通过充分定义已命名好的关系集合,将包含相同成分或剂型的不同术语类型联系起来。这些关系包括has_ingredient/ingredient_of、constitutes/consists_of等,它们是双向的,且术语类型之间的关系是相对固定的。

通过关系集合,包含相同成分或剂型的各种术语类型构成了彼此连接、格式固定的语义网(Semantic Network)。图2显示了不同抽象水平西替利嗪(Cetirizine)的语义网。语义网的构建使得从药品成分的一种术语类型出发,可检索到该药品成分的其他各种术语类型信息。

RxNorm对包含同一种化学成分或剂量的各种术语类型构建了彼此关联的语义网,即从成分、通用名形式、商品名形式等任何入口均可查找到其他表达形式,通过关系传递实现不同表达水平、同种化学成分的药品信息交互。

3.4 RxNorm数据存储方式是支持多系统互操作的重要基础

RxNorm采用与UMLS相同的RRF格式文档进行信息存储。12个文档将不同来源表的药品信息充分保存,包括概念、属性、关系、变更信息等。其中,有三个文档记录弃用药品名称(OCN)信息,另有三个历史文档记录原词和RXNCUI的变更信息:RXNCUI、RXNATOMARCHIVE和RXNCUICHANGES,使得RxNorm收录的来源表原词和RxNorm规范术语集经历版本变化依然有完整的信息保存^[7]。

以退役(retired)的RXNCUI为例^[7]。当RxNorm发现两个概念同义并将其合并,将先发布的RxNCUI退休,并将其映射到后一个RxNCUI,在RXNCUI.RRF文档表示形式为:

CUI1 (退役)	106107
CUI1首次出现的版本信息	RXNORM_04AC_050210F
CUI2末次出现的版本信息	RXNORM_08AB_090302F
CUI1迁移或拆分的数量	1
CUI2 (退役CUI的同义概念)	844308

当在应用程序RxNav^[9]中检索RxNCUI=106107时,系统自动提示“106107”已废弃,并可映射到RxNCUI=844308(如图3所示)。

RxNorm在数据存储方面体现了对历史数据的完好保存,RxNorm不会因为源词表以及自身

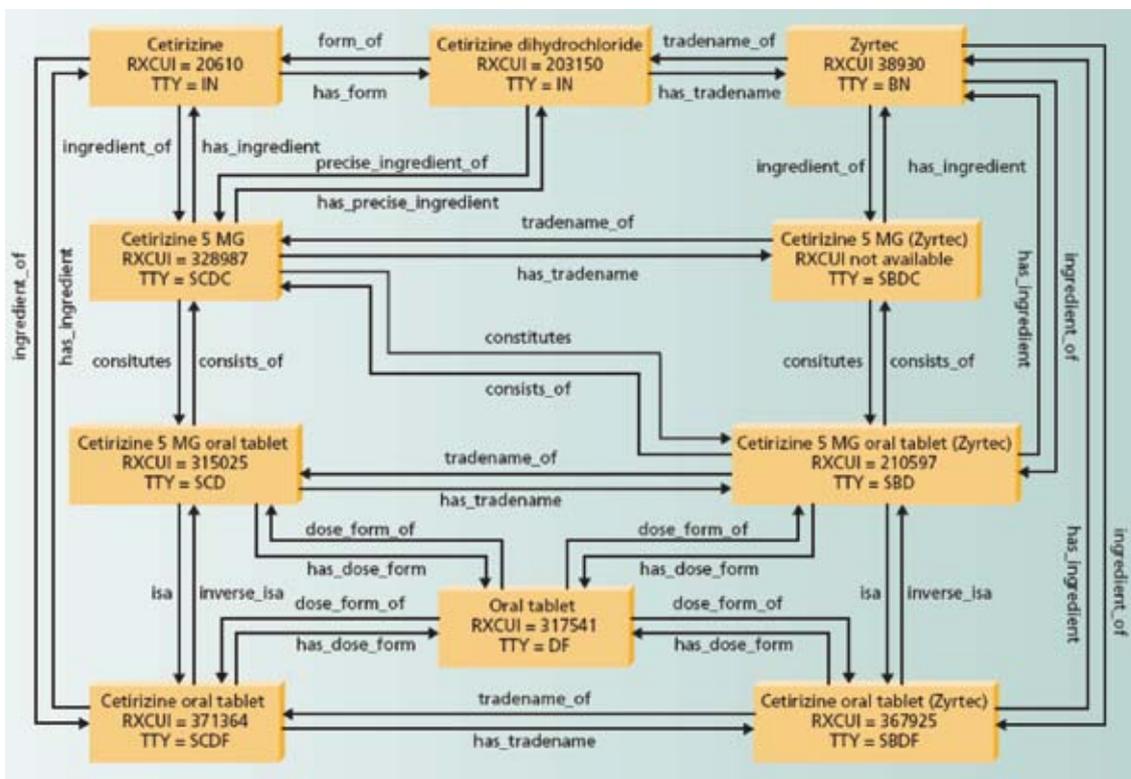


图2 连接不同抽象水平“西替利嗪 (Cetirizine)”的语义网

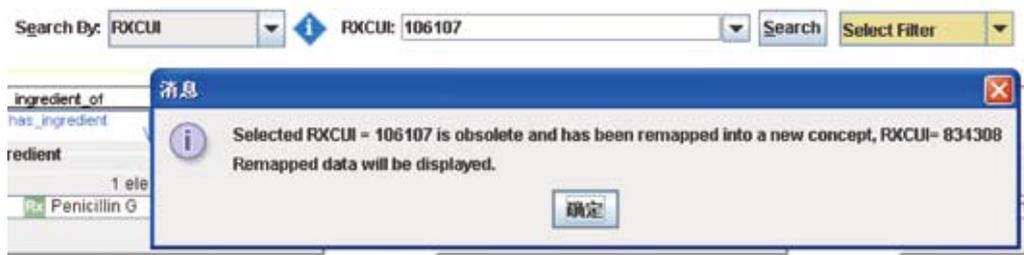


图3 RxNav系统退役CUI“106107”映射提示信息

的版本更迭造成数据的丢失，新旧数据的无缝映射使得系统之间能克服版本的障碍顺利地进行名称和编码的交互。

4 RxNorm支持系统互操作应用实例

4.1 用于国防部 (DoD) 和退伍军健康管理局 (VA) 之间交换数据^[10,11]

美国联邦政府通过多机构联

合努力达到卫生信息系统的互操作，采用RxNorm作为术语表药品标准来交换联邦机构药房数据信息。Parrish F等通过完成一个互操作标准的过程，将VUID (VA唯一标识符) 和NCID (DoD使用的概念ID) 映射到RxCUI，实现了国防部和退伍军人事务处之间实时的、双向编码的药品数据交换。

4.2 在临床数据仓库中管理医学词汇更新，促进多系统信息交互^[12]

为将两个医院药房信息系统的药品数据整合，斯坦福转化研究整合数据环境 (STRIDE) 项目采用RxNorm药品模型，将药品处方映射到RxNorm。为了保持数据的关联和及时更新，该系统将处方内容更新到RxNorm，同时保留原始意思和遗留数据的映射。RxNorm提供了丰富的内容和有效的药品表示模型，显著性地降低系统整合和维护的难度。

4.3 处理大型医疗药品数据,支持临床转化研究^[13]

在CTSA(临床转化科学奖)项目资助下,美国学术性医疗中心将临床处方数据存储到一个整合的数据仓库(Integrated Data Repository, IDR)中。然而各地处方集的数据编码在各参与机构中有显著性的不同,使得机构间药品数据查询难以采用同一标准,跨机构处方数据规范化的需求为实现整合大型、分布式数据集的目标提出了巨大挑战。

Wynden R等开发了健康本体映射器(Health Ontology Mapper, HOM),创造性地解决了这一问题。该系统在底层与NLM提供的RxNav术语服务器(RxNav terminology server)交流。通过RxNav的中间传导,任何基于i2b2.org的IDR处方数据都可自动翻译成RxNorm标准,各地的编码数据能映射到标准的术语集,而不需要再对来源的界面代码或数据模型进行转换。输入医疗机构的药品名称或NDC代码,可得到用户相关的、来自RxNorm的标准化药品名称和编码。

5 启示

RxNorm作为临床药品标准术语集支持多系统互操作,其成功经验值得我们思考。

5.1 系统合理设计与数据规范化处理

RxNorm借鉴UMLS多词表整合思路,做到以简单模式融合多个异构来源表的丰富信息^[5];同义与歧义概念的严格界定,保障数据无歧义、安全的交互;选用处方形式{成分+剂型+剂量[商品名]}为概念表达模式,互操作结果能更好满足临床需求;语义规范化概念名称、RxCUI将不同形式、不同编码的同义概念连接起来,构成术语互操作的枢纽。同时,高度的规范化也要求RxNorm数据处理的关键环节需要相当程度的人工干预^[2],可见细粒度规范命名表与高效语义互操作支持工具的构建,除了先进的知识组织方式,计算机辅助的人工判断依然不可或缺。良好的运行机制、协作的建设团队以及构建联邦级药品标准名称表的使命感是保障大规模药品术语规范化工作高效进行的关键。

5.2 数据更新与服务创新,激发词表生命力

对知识组织系统来说,更新与维护至关重要。RxNorm每月更新、每周补充变化数据,及时捕捉美国药品市场和处方信息的动态,保持数据收录的全面性与时效性。为促进实际应用,RxNorm推出多个公众应用终端,包括RxNav、

RxTerm、MyMedicationlist与MyRxPad等,并着手致力于电子处方子集的开发^[3],使得其应用范围不断深入、个性化。广泛的应用需求为其更好的支持系统间互操作、促进临床决策提供源源不断的推动力。

6 结语

在完善的数据保存方式支撑下,RxNorm基于语义的规范化概念名称、唯一概念标识符RxCUI、丰富与规范化的编码属性以及多层次术语类型与语义网是实现多词表语义互操作的机制。它科学的组织方式、规范的加工模式和丰富的后台操作使其成为支持美国大型药品信息系统互操作的关键工具,有力的促进了卫生数据的广泛共享与高效交互^[3]。在卫生信息化深入发展的背景下,如何借鉴RxNorm先进经验来构建适合我国国情的药学词表标准,促进异构信息系统的数据交互,满足医务人员和公众多目的健康信息需求,值得做进一步的研究与探索。

致谢

本文得到了李丹亚、胡铁军、李军莲、钱庆和李姣等多位老师宝贵的指导意见,在此一并表示感谢。

参考文献

- [1] LIU S, MA W, MOORE R, et al. RxNorm: Prescription for Electronic Drug Information Exchange [J]. IT Professional, 2005(9/10): 17-23.
- [2] RxNorm overview [EB/OL]. [2012-04-01]. <http://www.nlm.nih.gov/research/umls/rxnorm/overview.html#>.
- [3] NELSON S J, ZENG K, KILBOURNE J, et al. Normalized names for clinical drugs: RxNorm at 6 years [J]. Med Inform Assoc, 2011(18): 441-448.
- [4] 2011AB RxNorm Source Information [EB/OL]. [2012-04-01]. <http://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/RXNORM/index.html>.
- [5] 李丹亚,胡铁军,李亚子,等.UMLS多词表整合机制研究[G]//中国医学科学院医学信息研究所2011年学术年会论文集:15-24.
- [6] 2011AB RxNorm Source Information-Statistics and Sample Data-Source Overlap [EB/OL]. [2012-05-17]. <http://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/RXNORM/overlap.html>.
- [7] RxNorm Technical Documentation Version 2012-2 [EB/OL]. [2012-04-02]. http://www.nlm.nih.gov/research/umls/rxnorm/docs/2012/rxnorm_doco_full_2012-2.html.
- [8] DOLIN R H, ALSCHULER L. Approaching semantic interoperability in Health Level Seven [J]. J Am Med Inform Assoc. 2011, 18(1): 99-103.
- [9] RxNav Java应用界面[CP]. [2012-06-29].
- [10] PARRISH F, DO N, BOUHADDOU O, et al. Implementation of RxNorm as a terminology mediation standard for exchanging pharmacy medication between federal agencies [C]// AMIA AnnuSympProc, 2006: 1057.
- [11] BOUHADDOU O, WARNEKAR P, PARRISH F, et al. Exchange of computable patient data between the Department of Veterans Affairs (VA) and the Department of Defense (DoD): terminology mediation strategy [J]. J Am Med Inform Assoc, 2008(15): 174-83.
- [12] PODCHYNSKA T, HERNANDEZ P, FERRIS T, et al. Managing Medical Vocabulary Updates in a Clinical Data Warehouse: An RxNorm Case Study [C]// AMIA AnnuSympProc, 2010-11-13, 2010: 477-81.
- [13] WYNDEN R, ANDERSON N, CASALE M, et al. Using RxNorm for cross-institutional formulary data normalization within a distributed grid-computing environment [C]// AMIA AnnuSympProc, 2011, 2011: 1559-63.

作者简介

李芳 (1980-), 硕士, 研究方向: 生物医学知识组织方法与系统建设。E-mail:helenlifang2004@163.com

Mechanism of RxNorm in Supporting Multi-Terminology Semantic Interoperation

Li Fang, Chen Ying, Hou Li, Ji Yujing / Institute of Medical Information, Chinese Academy of Medical Sciences, Beijing, 10002

Abstract: Semantic interoperability among different information systems is a major concern in the development of knowledge organization systems within the Web circumstance. In the field of medication, RxNorm is a representative tool to support semantic interoperation between different drug systems. Based on systematic analysis of knowledge organization pattern, the article reveals the mechanism of RxNorm in supporting multi-terminology semantic interoperation, from the aspects of its concepts representation, attributes normalization, structured semantic network as well as data preservation files. Furthermore, typical application examples of RxNorm are listed and some thoughts about RxNorm are presented.

Keywords: RxNorm, Semantic Interoperation, Drug Terminology, RxCUI

(收稿日期: 2012-08-01)