

本文引用:朱玲,李敬华,于彤,于琦,田野,唐旭东.基于本体的功能性胃肠病知识库构建研究[J].湖南中医药大学学报,2018,38(10):1154-1160.

基于本体的功能性胃肠病知识库构建研究

朱玲¹,李敬华¹,于彤¹,于琦¹,田野¹,唐旭东^{2*}

(1.中国中医科学院中医药信息研究所,北京 100700;2.中国中医科学院西苑医院,北京 100200)

[摘要] 基于 Web 本体语言 (web ontology language, OWL) 构建功能性胃肠病 (functional gastrointestinal disorders, FGIDs) 的领域本体,在此基础上扩充海量实例,建立包含本体库、规则库、知识库的 FGIDs 本体知识库模型,实现基于语义网规则语言 (semantic web rule language, SWRL) 的诊断、治疗等规则的表示及推理。此外,还通过文献计量学对 FGIDs 领域的主题词分布、医家、方剂、中药等进行了统计分析及相关规则计算,提示四君子汤和六君子汤是领域中使用最为广泛的方剂。通过对 FGIDs 领域知识的揭示,语义关联的表达,为建立可共享、可复用的中医疾病领域本体知识库提供一定参考。

[关键词] 功能性胃肠病;本体;知识库;语义网规则语言

[中图分类号] R256.3 **[文献标志码]** A **[文章编号]** doi:10.3969/j.issn.1674-070X.2018.10.015

Construction of Knowledge Base for Functional Gastrointestinal Diseases Based on Ontology

ZHU Ling¹, LI Jinghua¹, YU Tong¹, YU Qi¹, TIAN Ye¹, TANG Xudong^{2*}

(1. Information Institute on Traditional Chinese Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China;

2. Xiyuan Hospital, China Academy of Chinese Medical Sciences, Beijing 100200, China)

[Abstract] The domain ontology of "functional gastrointestinal diseases (FGIDs)" has been constructed by Web Ontology Language. On this basis, we expand massive instances and establish a knowledge base model for FGIDs ontology containing ontology base, rule base, and knowledge sub-base. The knowledge base model implements the representation and reasoning of rules such as diagnosis and treatment based on semantic web rule language. In addition, through the bibliometric, the distribution of the subject words in the field of FGIDs, doctors, prescriptions, traditional Chinese medicine (TCM), etc., were statistically analyzed and calculated based on association rule, suggesting that Sijunzi Decoction and Liujunzi Decoction are the most widely used prescriptions in the field. In summary, this paper provides a certain reference for the establishment of a shareable and reusable domain ontology knowledge base for TCM diseases through the disclosure of knowledge in the field of FGIDs and the expression of semantic association.

[Keywords] functional gastrointestinal diseases; ontology; knowledge base; semantic web rule language

功能性胃肠病 (functional gastrointestinal disorders, FGIDs) 是指具有胃肠道症候群,但缺乏器质性疾病或能解释其临床表现依据的一组疾病^[1]。罗马 III 标准将成人 FGIDs 分为 6 大类 28 种疾病^[2],为其诊断治疗提供了新的权威标准。国外文献报道功能

性消化不良 (functional dyspepsia, FD) 患病率为 20%~40%^[3],国内报道各种功能性肠病总患病率在 43%~46%^[4]。其不但发病率高,而且重叠症状多^[5],个体差异大,长期治疗效果欠佳,严重影响患者的生活质量。同时 FGIDs 的诊治费用比较高,给社会及个体

[收稿日期] 2018-05-28

[基金项目] 中国中医科学院“十二五”重点领域研究专项资助 (ZZ0808002)。

[作者简介] 朱玲,女,副研究员,博士,研究方向:中医药领域本体建设以及现代信息技术在中医药领域的应用。

[通讯作者] * 唐旭东,男,博士研究生导师, E-mail: txdy@sina.com。

带来沉重的经济负担。近年来西医对 FGIDs 的研究广泛而深入,但中医对 FGIDs 的治疗则缺乏规范、严格、合理的临床实验研究,没有形成业内公认可以共享的知识体系,后学者对此不免心生困惑。本体^[6]作为一种业界公认的知识建模工具,其支持对概念层级结构的表示,并且对概念与概念之间的关系有很好的描述,为解决概念理解的歧义提供了可能,同时支持逻辑推理,为知识的共享提供了基础。

本文使用斯坦福大学的开源本体构建工具 protégé 5.1 建立 FGIDs 本体库,对该领域内的概念以及概念之间的约束和关系进行描述,利用语义网规则语言(semantic web rule language, SWRL)建立 FGIDs 的规则库,初步实现 FGIDs 的诊断及治疗推理^[7]。同时,以领域本体为基础,检索相关文献,构建领域知识库,为未来领域本体实例的扩充奠定基础。

1 FGIDs 本体的建立

1.1 确定领域本体目的及范围

FGIDs 本体主要包含:事实知识(如 FGIDs 的症状表现)、经验性知识(如病因与症状之间的关系,方剂或药物和症状之间的关系)和描述病理过程的知识等。

1.2 获取领域中的重要知识

首先,从中文医学主题词表、中医药语言系统^[8]、内科教材、辞典、相关标准、相关专科书籍及期刊文献中筛选出重要专业术语,根据该研究目的和范围收集所有专家认可的概念、语义、属性、实例等,经过梳理建立术语表,并将术语的定义一并收入,力求术语集的完整性及准确性。重要术语见表 1。

表 1 FGIDs 重要术语表

疾病	症状	病因	医家	实验室检查	药物
儿童/青少年 FGIDs	腹痛	体质	王辉武	X 线检查	香砂六君丸
新生儿/婴幼儿 FGIDs	恶心	劳倦	唐江山	内脏感知监测	健胃消食口服液
功能性肛门直肠病	呕吐	外感	何世东	消化道压力测定	参苓白术颗粒
功能性肠病	便秘	微生物	单兆伟	胃电图	复方陈香胃片
功能性胃十二指肠病	腹泻	-	单秋华	血常规	越鞠丸
功能性腹痛综合症	头昏	-	周仲瑛	食管	H ₂ 受体拮抗剂
功能性食管病	抑郁	-	-	胃腔内 pH 测定	抗抑郁焦虑类药
胆囊和 Oddi 括约肌功能障碍	腹痛	-	-	-	人参健脾丸
儿童/青少年 FGIDs	恶心	-	-	-	保济口服液

1.3 确定基本概念及其类别、关系阶段

上一步骤中已经产生了大量的概念,按照一定的规则自上而下将其分组,形成不同的类别。本文以“FGIDs”为目标,参考中医药学语言系统^[8],利用斯坦福大学开发的本体构建工具 Protégé5.1 将两个概念之间的语义关系进行清晰、准确的揭示与定义(见表 2),进而建立领域框架知识,这有助于概念间语义的挖掘,同时也为未来的自动推理提供了可能。

表 2 所示为“功能性胃肠疾病”直接涉及的属性。其他对象属性(object attribute)还包括 is a(是……), instance of (……的实例), attribute of (……的属性), result in (导致), be good at (擅长), consist of (由……组成), prevent(预防)等,其中导致、组成、预防、现象表达、治疗等属于 inverse functional(互为反义的一组关系)。其中 be good at(擅长)主要用于表示医家擅长于某个学科领域,这对于临床就医推荐具有一定的实用价值。

下一步定义 FGIDs 本体中的实例,并将这些实例存入 FGIDs 本体库中,为推理做准备。通过以上几个步骤构建完成的 FGIDs 本体模型。见图 1。

表 2 FGIDs 关键术语的语义关系一览表

概念词	语义关系	相关概念
FGIDs	与……同时发生	自主神经紊乱
	位于	脾、胃
	位于	大肠、小肠
	使用	消化道压力测定
	在……之前	进餐
	在……之后	进餐
	有……现象表达	肝胃不和
	有……现象表达	腹胀、腹泻、恶心、呕吐
	被……治疗	天枢穴
	被……治疗	四磨汤
	被……导致	心理因素

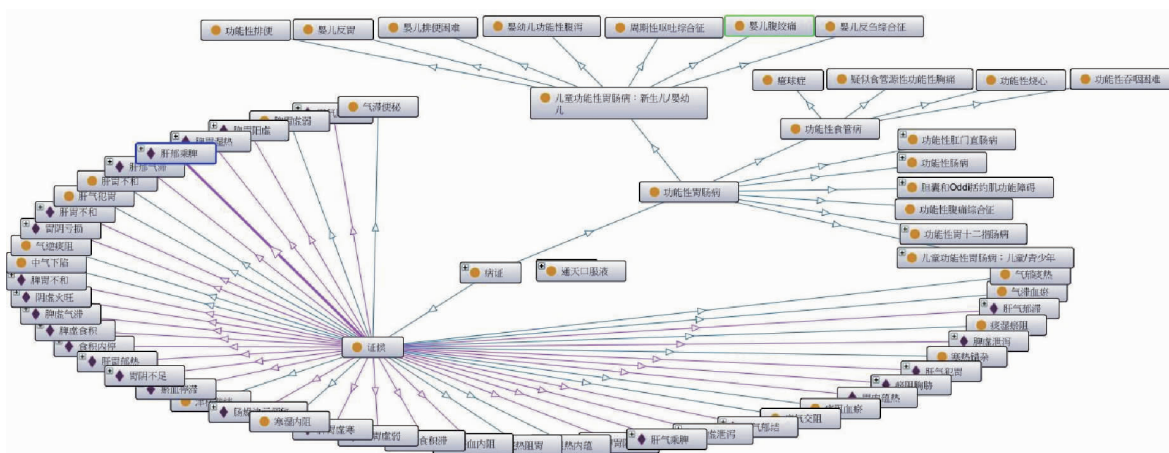


图 1 FGIDs 本体知识库的框架

1.4 FGIDs 本体展示

FGIDs 本体共有类别 178 个, 对象属性 14 个, 实例 481 个, 能够实现对 FGIDs 的概念层级分类以及概念间的语义关系展示。其中疾病部分按照最新的罗马 III 疾病分类表示, 所有左上角有加号的, 都表示有下位类存在。见图 2。

如寒湿内阻证的临床表现为神疲、四肢乏力、脉沉迟, 治法为温中散寒, 化湿和胃, 常用方剂为桂附理中汤。见图 3。

2 FGIDs 本体知识库的构建

2.1 FGIDs 本体知识库模型

知识库是面向组织的、集成的、随时间变化的数

据存储空间, 是事实、规则和概念的集合。与数据库相比, 知识库除了包含数据库存储的事实, 还包含数据库无法存储的常识、经验、规则等启发性知识^[9], 这是二者的根本区别。其基本内容可以包括表示领域中概念层次和属性关系描述的实例集。R 为规则集, 表示对实例知识进行推理的启发式规则的集合。FGIDs 知识库模型(如图 1)采用本体知识表示的方式, 以本体作为 FGIDs 的领域概念、概念间关系以及概念间内在关系的形式化表达, 进一步利用本体概念和关系进行 FGIDs 诊断和治疗规则的组织, 完成规则库的构建。

传统建立分类概念的层次结构有三种方法^[10], 自下而上法, 自上而下法和综合法。本研究拟采用综

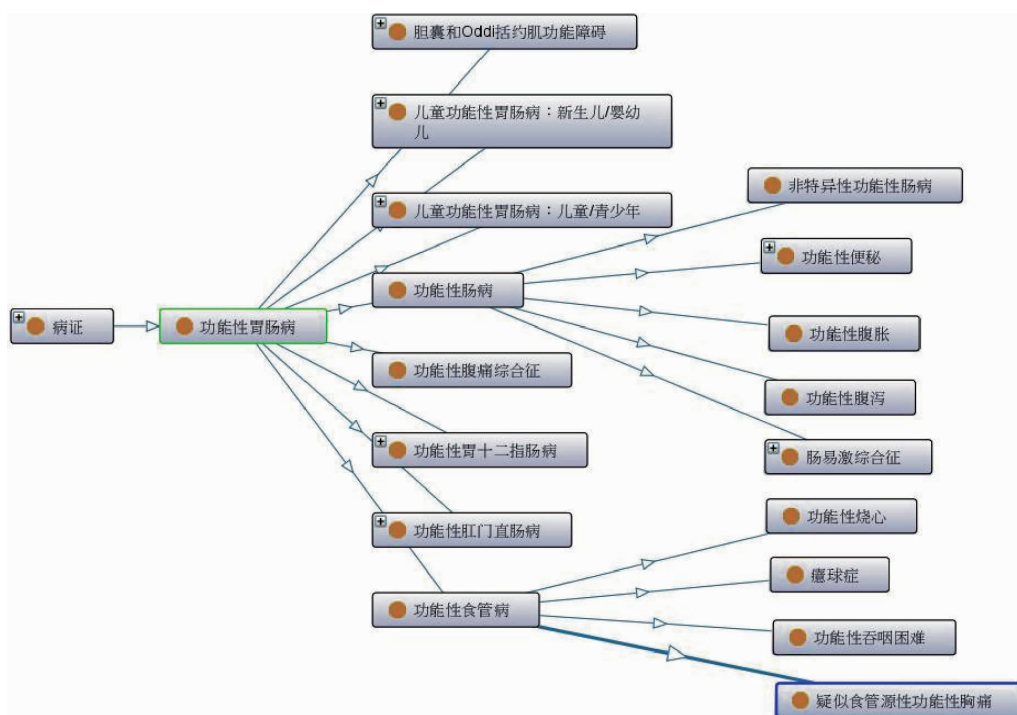


图 2 罗马 III 疾病分类表

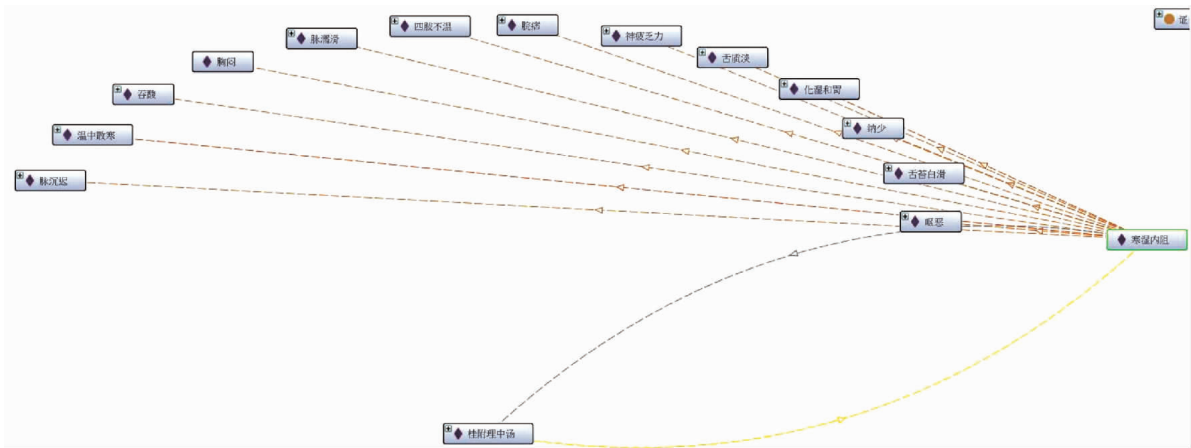


图 3 寒湿内阻的本体表达

合法,参考中医临床术语系统^[11]的顶层架构,也考虑领域自身的特点。因此,FGIDs 本体知识库模型包含本体库、规则库、知识子库三个子库,其中知识库主要分为以下几个大类:病证、症状体征、病因病机、治则治法、病人、药物、中医操作、实验室检查、器械设备、医家、机体形态。见图 4。

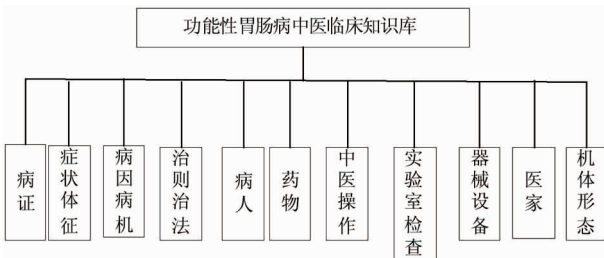


图 4 FGIDs 本体知识库架构

2.2 基于本体的 FGIDs 知识库

领域本体概念层构建完成之后,概念的体系结构相对比较稳定,但领域本体知识库的质量不仅限于概念层的科学性和丰富程度,还在于实例部分的可扩展性,可维护性等因素,这也是本体构建的重点之一,同时也决定了领域本体作用能否得到更好的发挥。

本文在 FGIDs 领域本体的基础上进行实例的扩展,试图构建领域知识库。首先以“消化不良、肠易激综合征、结肠假性梗阻、神经源性肠道功能障碍、功能性结肠疾病、结肠疾病、功能性”等为主题词检索中国知网、维普、万方、中医药数据库检索系统、PubMed、中医药知识服务平台^[12]等,全面收集中医药学、针灸、气功、按摩、保健等方面的数据,形成总篇数约 1.2 万的包含中文期刊、英文期刊、古籍、科普 4 大类文献资源。在此基础上,以领域本体为架构,对其进行包括主题标引、分类、知识抽取、搭建数

据库、平台服务等操作,形成包括 ID、中文题名、英文题名等 22 个字段的 FGIDs 知识库。

本文选用 SQL Server 关系数据库作为构建知识库的基础,存储行业信息化的 14 大类内容实例,即实例数据库;文档库存有知识库系统中各类知识的附件信息,是供用户下载的文档内容,文档容量比较大,整个知识库包含 14 个子库,总量为 25 704 条知识记录。见图 5。

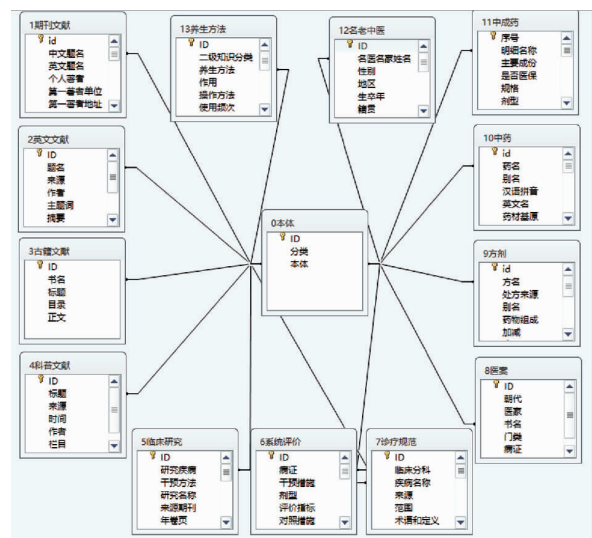


图 5 知识库架构

2.2.1 主题词分布 共有 30 176 个主题词,其中频数分布前 47 项如表 3 所示。FGIDs 主要涉及疾病为消化不良、肠易激综合征、功能性结肠疾病,症状有便秘、腹泻等,治疗中药有柴胡、白芍、党参,处方以痛泻要方等为主。其中柴胡疏肝散、逍遥散作为疏肝解郁的经典方剂,在治疗 FGIDs 的方剂中分别排名第 5、6,仅次于痛泻要方、四逆散、半夏泻心汤、参苓白术散,充分说明了情志因素是不可忽视的病因。

表 3 FGIDs 主题词分布一览表

序号	主题词	频数	序号	主题词	频数
1	消化不良	3 011	24	四逆散	156
2	肠易激综合征	2 448	25	中成药	151
3	便秘	1 114	26	度哌酮	146
4	腹泻	858	27	半夏泻心汤	141
5	白术	580	28	参苓白术散	138
6	柴胡	556	29	柴胡疏肝散	132
7	结肠疾病,功能性	467	30	穴,中脘	128
8	白芍	417	31	慢性病	120
9	党参	381	32	逍遥散	120
10	痛泻要方	355	33	厚朴	116
11	呕吐	325	34	促胃动素	116
12	恶心	285	35	名医经验	114
13	茯苓	276	36	香附	113
14	辨证分型	271	37	大黄	107
15	穴,足三里	235	38	胃排空	105
16	方剂化裁	227	39	黄连	103
17	治疗结果	211	40	肝郁脾虚	98
18	黄芪	198	41	胃肠活动	94
19	陈皮	186	42	化疗反应	94
20	枳壳	177	43	西沙必利	93
21	枳实	166	44	老年人	91
22	木香	166	45	穴,内关	89
23	穴,天枢	160	46	疏肝	88

2.2.2 临床研究 共有 1 575 项 FGIDs 的临床研究论文发表,对其研究内容概要分析如表 4 所示。采用的干预措施还是以中西医结合为多,这与上文提及的发表杂志的选择不谋而合。半夏泻心汤作为寒热平调之剂,具有消痞散结之功效,其主症与 FGIDs 诸多吻合,故在临床研究中也是被采用最多的方剂。

如表 5 所示,FGIDs 采用的研究方法还是以单纯的临床试验为多,随机对照试验不足其十分之一,当然这应该是中医临床研究的普遍问题,也是亟需重视与提高的问题。

2.2.3 医案 检索中国中医科学院中医药信息研究所研制的中医医案系统^[13],共检出脾胃病、腹泻、便秘等相关医案 1 477 则,其出现较多的医家有叶天士、张聿青、薛立斋等,周仲瑛作为现代医家排名第 5,一方面说明其治疗 FGIDs 医案数量多,二则说明其在脾胃病治疗方面有独到之处。见表 6。

2.2.4 方剂 知识库中总计共有 7 068 个方剂。其中使用频次 ≥ 450 的中药见表 7。甘草作为调和诸药之药老排名第 1,当无疑义,也无意义。人参、白术、陈皮、木香,或补益脾气,或疏理脾气,都是临床

表 4 FGIDs 主要临床研究文献分析

序号	干预	篇数
1	中西医结合	103
2	自拟方	50
3	辨证论治	35
4	半夏泻心汤	32
5	四逆散	27
6	六味安消胶囊	25
7	香砂六君汤	22
8	四磨汤	20
9	柴胡疏肝散	19
10	逍遥散	16
11	针刺	15
12	中药	13
13	枳实消痞汤	10
14	耳穴贴压	10
15	六味能消胶囊	9
16	推拿	8
17	疏肝健脾法	8
18	藿香正气液	8
19	穴位埋线	7
20	疏肝和胃汤	7

表 5 FGIDs 临床研究方法分析

研究方法	篇数
临床试验	1 362
随机对照试验	102
病例报告	70
对照临床试验	22
对比研究	19

脾胃科常用之中药。

将 7 068 首方剂中的药物录入 Microsoft Excel 数据库,把涉及的每一种药物作为一个变量,把一个方剂作为一条记录,根据方剂中药物的出现与否按 1、0 分别对其进行赋值。数据录完进行数据整理,锁定备用。统计方法采用 SPSS 17.0 和 SPSS Clementine11.1。频率分析采用 SPSS 17.0 中的频率描述统计方法,药物关系模型的建立及药物配伍规则的提取采用 SPSS Clementine 关联规则中的广义规则归纳(generalized rule induction,GRI)算法。关联规则是一种数据挖掘的方法,它能够发现数据库中不同变量之间的联系规则,确定数据中不同领域之间的关系。中药之间配伍关系如表 8 所示,置信度排名前 10,均为四君子汤或六君子汤的配伍,可见其在临床上广泛使用。

2.2.5 名医 知识库中包含治疗 FGIDs 的国家级和省级名医(王辉武、唐江山、何世东、单兆伟、单秋

表 6 FGIDs 医案数量排名 ≥20 的医家

序号	医家	数量
1	叶天士	325
2	张聿青	65
3	薛立斋	59
4	也是山人	38
5	周仲瑛	37
6	朱丹溪	32
7	熊惠生	26
8	丁甘仁	23
9	张镜人	22
10	邵兰荪	22
11	赵绍琴	21
12	朱良春	20
13	李振华	19
14	徐景藩	17
15	李士材	17
16	程杏轩	17
17	赵守真	16
18	张子和	16
19	王孟英	15
20	汪石山	15

表 7 FGIDs 使用频次 ≥450 的中药一览表

序号	中药	频次	百分比/%
1	甘草	2 447	4.24
2	人参	1 534	2.66
3	白术	1 492	2.59
4	陈皮	1 397	2.42
5	木香	1 097	1.90
6	当归	1 015	1.76
7	厚朴	903	1.57
8	干姜	867	1.50
9	茯苓	833	1.44
10	半夏	760	1.32
11	枳壳	628	1.09
12	肉桂	608	1.05
13	黄连	553	0.96
14	槟榔	544	0.94
15	丁香	538	0.93
16	附子	533	0.92
17	黄芩	512	0.89
18	白芍	480	0.83
19	大黄	450	0.78

华、周福生、高金亮、谢晶日、甘爱萍、任明、汪玉琴、陈永灿、单秋华、路广晁)共 14 位,在排名前 10 的专家中广东占据了 3 席,与平丽等^[4]进行的广州地区 FGIDs 患病情况调查不谋而合,其研究表明,FGIDs 是广州地区的常见病、多发病。

表 8 FGIDs 方剂中配伍关系关联规则

规则前项	规则后项	置信度	支持度
人参	甘草	0.13	0.07
陈皮	甘草	0.13	0.06
茯苓	甘草	0.13	0.06
白术	甘草	0.12	0.06
白术	人参	0.11	0.06
人参	白术	0.11	0.06
白术	茯苓	0.11	0.05
茯苓	白术	0.11	0.05
茯苓	人参	0.1	0.05
人参	茯苓	0.1	0.05
白术	陈皮	0.1	0.05
陈皮	白术	0.1	0.05
人参	陈皮	0.09	0.05
厚朴	甘草	0.07	0.06
厚朴	陈皮	0.07	0.05
半夏	甘草	0.06	0.06
干姜	甘草	0.06	0.05
半夏	陈皮	0.06	0.05
半夏	茯苓	0.06	0.05

注:置信度:0.05 支持度:0.05。

3 FGIDs 诊断规则的建立及语义推理的实现

近年来,本体和规则的融合成为新的研究热点,SWRL、SQWRL 和 DL rule 等基于 OWL 或 DL 的新规则语言的出现,使 DL 规则成为语义描述和推理的最佳选择之一。SWRL 是由 OWL 子语言 OWL DL 与 OWL Lite,以及 Unary/Binary Datalog RuleML 为基础的规则描述语言,其目的是为了与 OWL 知识库产生结合,提升 OWL 逻辑表达的能力,进而提高本体的推理能力。Java 专家 Shell 系统(Java Expert Shell System, JESS)是一种基于 Java 语言的 CLIPS 推理机,是基于产生式的向前推理引擎。其优点为简单实用及集成性高。

在 FGIDs 本体的基础上,根据临床指南、医学书籍和文献以及专家知识,在 The SWRL Editor 上编辑 FGIDs 的诊断及治疗规则。用 SWRL 语言表示部分规则如下:

```
病人(? x) ∧ 症状(? a)
>18 岁</age>
∧ show(? a, 上腹痛) ∧ show(? a, 焦虑)
→has(? x, 实验室检查阴性)
→has no association with(? x, 进食)
→has no association with(? x, 月经)
```

→has no complication

患者 x, 如果有腹痛、焦虑症状, 但实验室检查未见明显阳性, 就可以初步诊断患者是 FGIDs 之功能性腹痛综合征。

JESS 推理引擎主要由事实库、规则库以及推理机 3 部分组成, 首先基于建成的领域本体, 创建实例, 推理机将领域本体转化为 JESS 事实库, 再将 SWRL 规则转化为规则库, 推理机结合 JESS 事实库与规则库, 实现基于本体的语义推理, 实现对 FGIDs 领域潜在知识的揭示, 实现对疾病的早期诊断、查询以及药物推荐等服务。

肠易激综合征的症状描述为腹痛、腹泻、恶心欲吐、性情急躁, 用 SWRL 翻译, 就可以得到相应的规则, 腹痛 \wedge 腹泻 \wedge 恶心欲吐 \wedge 性情急躁 \rightarrow 肠易激综合征。

如患者张三(化名), 刻下症为腹痛、痛后即泻、泻后痛减、与情绪变化相关, 性情急躁、善太息、两胁胀满、纳差、恶心、舌淡红苔薄白、脉细弦, 最终诊断为肠易激综合征(腹泻型), 泄泻(肝郁脾虚)推荐口服奥替溴铵片, 口服马来酸曲美布汀片, 中药为痛泻要方。

4 小结

本文利用 OWL 对“FGIDs”领域中的概念、属性及概念和属性之间的关系进行知识表示, 将语义特征引入领域知识库, 建立起适合的知识体系框架, 从而实现了知识与知识之间的关联, 提炼出一个具有内在逻辑结构的 FGIDs 的中医临床知识体系, 实现了简单的推理, 为未来中医药在其他常见病治疗使用过程提供全面、可共享的知识库构建范式, 为中医临床工作者全方位学习、查询中医在 FGIDs 的古今应用实例提供了便利的检索功能, 使之能更好地把握 FGIDs 的各种属性的内在联系, 并且通过对知识库主题词、方剂等重要内容的文献计量学以及关联规则分析, 有利于进一步审视当前中医药治疗

FGIDs 的现状, 同时也对临床实践活动起到一定的指导和借鉴作用。

但基于时间限制, 本文未对基于本体的知识库做进一步的验证研究, 如利用本体查询语言和本体推理机, 结合具体应用场景实现基于语义的知识检索和更为复杂的推理等, 这应该是未来的研究方向之一。

参考文献:

- [1] 柯美云. 功能性胃肠病并非单纯功能病[J]. 中国实用内科杂志, 2006, 26(10):721-722.
- [2] DROSSMAN D A. 罗马 III: FGIDs[M]. 3 版. 北京: 科学出版社, 2008:5.
- [3] 李瑜元. 功能性胃肠道病的精神心理因素及治疗[J]. 国外医学内科学分册, 1998, 25(2):105-107.
- [4] DROSSMAN D A, LI Z, ANDRUZZI E, et al. U. S. Household survey of functional gastrointestinal disorders[J]. Digestive Diseases & Sciences, 1993, 38(9): 1569-1580.
- [5] 王捷虹, 靳光荣, 卞立群, 等. 唐旭东教授对功能性胃肠病症状重叠现象的学术观点及其中医治疗经验[J]. 四川中医, 2016, 34(3):8-9.
- [6] 杨月华, 杜军平, 平 源. 基于本体的智能信息检索系统[J]. 软件学报, 2015, 26(7):1675-1687.
- [7] 朱 玲, 李敬华, 于 琦, 等. 中医哮喘领域本体的构建[J]. 中国实验方剂学杂志, 2017, 23(15):222-226.
- [8] 于 彤, 崔 蒙, 李海燕, 等. 中医药学语言系统的语义网络框架: 一个面向中医药领域的规范化顶层本体[J]. 中国数字医学, 2014, 9(1):44-47.
- [9] 赵秀翔, 白晓丽. 基于本体的 FMEA 知识库的构建方法研究[J]. 武汉理工大学学报(信息与管理工程版), 2009, 31(1):87-90
- [10] 战学刚, 林鸿飞, 姚天顺. 中文文献的层次分类方法[J]. 中文信息学报, 1999, 13(6):20-25.
- [11] 中医临床术语 [EB/OL]. [2016-07-14] <http://192.168.200.83:8080/tpweb/MainInitAction>.
- [12] 中医药知识服务平台[EB/OL][2018年7月14日] <http://www.tcmkb.net>.
- [13] 于 琦, 王映辉, 李敬华, 等. 中医名医医案分析系统研究[J]. 中国数字医学, 2015, 10(3):51-53.
- [14] 平 丽, 李瑜元, 聂玉强, 等. 功能性胃肠病患病情况调查[J]. 实用医学杂志, 2003, 19(4):424-426.

(本文编辑 贺慧娥)