

语义关系在语言系统中的作用及现状分析*

□杨 阳 崔 蒙** 李园白

(中国中医科学院中医药信息研究所 北京 100700)

摘 要:本文分析了当前国内外4大语言系统的语义关系发展情况,并着重分析了中医药类语言系统语义关系的发展存在问题,最后针对中医药领域内语义关系脱离临床实践的发展问题,建议从真实的病历中验证及提取凸显临床特色的语义关系。

关键词:中医药 语言系统 语义关系 综述

中医药学绵延发展 5000 年,积累的数据可谓浩如烟海,如何有效地管理和利用这些数据,从当前以数据处理为主导的初级阶段过渡到以知识处理为主导的高级阶段,成为中医药信息研究人员关注的焦点。中医药语言系统的开发为中医药学术语的集成、电子病历知识库系统的构建、医疗数据的采集以及检索共享平台的建立提供了良好的发展契机。

一、语言系统

从功能上来说,语言系统是支持自然语言与受控语言链接和转换的高度专业化系统,它的理想状态应该是:在系统外部,用户可以灵活地使用自然语言检索和组织信息,不需要看见和直接使用受控语

言;在系统内部,存在着以超级知识库为基础的、高度专业化的受控语言,支持用户自然语言提问的转换^[1]。语言系统是知识工程的一种技术,它通过本体论途径建立,同时需要语言学的支撑^[2-3]。

二、语言系统中的语义关系

医学语言系统最基本的组成部分是术语表和语义系统,其中语义系统又可分为语义类型和语义关系,所谓语义类型,是概念所属的语义形态,语义关系则存在于抽象的类型或具体的概念之间,用来表达生物医学领域中概念之间的重要关系。语义类型和语义关系共同构成网状的语义结构,在这种网状的结构中,语义类型相当于语义网络的节点,而节点与节点之间的关系则是语义关系^[4-5]。

关联是人类思维的基本方式,是智能的表现,语

收稿日期:2009-07-24

修回日期:2009-08-09

* 国家“十一五”科技支撑计划项目“中医药诊疗与评价技术研究项目中医药临床效应评价研究”的子课题(2006BAI08B02-03):基于临床文献数据独立评价中医药临床效应的方法研究,负责人:曹洪欣,崔蒙;中国中医科学院基本科研业务费自主选题项目(Z02043):中药科学数据汇交的数据模式研究,负责人:杨阳。

** 联系人:杨阳,初级职称,中国中医科学院博士在读,主要研究方向:中医药信息研究方向, Tel:010-64048243, E-mail:sanyangkt@126.com;崔蒙,本刊编委,博士生导师,主要研究方向:中医药信息学的学科建设,中医药信息数据库与网络建设,中医药信息数据的挖掘与利用研究,以及中医药软科学战略研究, Tel: 010-64013395, E-mail:cui@mail.cintcm.ac.cn。

义系统通过语义关系在领域内可以严格规范概念之间的关联,电子病历通过语义关系可以将诊疗信息的生成方式由被动记录改变为主动获取,在语义检索时通过关系可以由被动获取知识改变为主动获取^[6]。

三、国内外医学语言系统的语义关系发展情况

在医学领域,依据不同的开发策略和服务目的,国内外开发了一些医学语言系统来支撑和促进医学信息化的发展,比较有代表性的有:美国病理学家学会开发的系统化医学术语集(SNOMED),美国国立医学图书馆开发的统一的医学语言系统(UMLS),以及中国中医科学院中医药信息研究所研制的中医药学语言系统(TCMLS)。

1. SNOMED 语义关系发展情况

SNOMED (Systemized Nomenclature of Medicine) 是一部在国际医疗服务中得到广泛应用的大型临床卫生保健术语集,它的普及和利用产生的是一个通用的临床数据库,能够实现临床信息的即时存取功能^[7]。SNOMED 由概念表、描述表和关系表组成,概念表收录了 366,170 个具有唯一含义并且经过逻辑定义的概念,描述表收录了大约 993,420 条同义术语说明^[8],关系表中大约有 146 万组语义关联,这些具有特定意义的语义关系的连接,使两个医学概念间能够关联在一起,形成数量极丰富的语义描述,使临床信息的应用具有极大的灵活性。

SNOMED 的语义关系词分布在语义类型 linkage concepts 中,关系表中的 146 万组语义关联信息,均通过语义关系词连接,从近几年的发展看,SNOMED 语义关系词调整力度较大,2003 版^[9]与 2002 版的语义关系词相比,虽然只是在总量上从 41 个增加到 42 个,但却是在去掉原有 11 个的基础上新增 12 个,变化的幅度虽大,但所有的语义关系词均是在语义类型 linkage concepts 范围内的波动。

2. UMLS 语义关系发展情况

(1) UMLS 语义关系基本情况。

UMLS 是美国国立医学图书馆自 1986 年开始研制的情报检索语言集成系统,主要由超级叙词表(Metathesaurus)、语义网络(Semantic Network)和专家词典(Specialist Lexicon)3 个部分组成。

UMLS 共有 54 种语义关系,这些关

系有些取自原有的索引表架构,有些是在建构 UMLS 时所赋予的^[10]。UMLS 将其归纳为等级关系和非等级关系,等级关系包括了“is_a”和“part_of”两个语义关联词,而另外的 52 种非等级语义关系归纳为“物理上相关”、“空间上相关”、“时间上相关”、“功能上相关”和“概念上相关”5 大类。这 5 类语义关系统领整个非等级关系,使其分类清晰结构明确。因此除了“conceptually_related_to”用了 2 次之外,另外的 4 个均为特定用词,只用来归类关系而没有连接具体的语义类型。

(2) UMLS 和 SNOMED 语义关系的对应情况。

根据调研发现 UMLS 的语义关系完全含纳或映射了 SNOMED2002 版以及 2003 版关联词,但由于两个系统应用角度的不同,以及应用时所指概念内涵和外延的差别,将 UMLS 的 54 个关联词和 SNOMED 的 41 个关联词相比,含义明显统一的关联词仅为 11 个(见表 1)。

3. TCMLS 语义关系发展情况

TCMLS 由中国中医科学院中医药信息研究所于 2001 研制开发,该系统由基础词库与语义系统两部分组成。它的语义关系采用了 UMLS 的分类形式,在采纳 UMLS 本身 54 个语义关联词基础上,根据中医药学科特点添加 4 种新语义关系,共形成 58 种语义关系。

4. 中医临床术语集语义关系发展现状

此部术语集由中国中医科学院中医药信息研究所于 2006 年研制,其语义关系词包含在语义类型“连接词”中,连接词又分为连接断定词、属性词、背景依赖性分类词、相关关系词、副主题词 5 大类。连接断定词是一些辅助实体词连接成句子的词,属性

表 1 SNOMED、UMLS 语义关系对应表

SNOMED	UMLS	意义
is_a	is_a	是
part_of	part_of	……的部分
component	consists_of	由……组成
has_active_ingredient	ingredient_of	……的成份,具有活性成份
occurrence	co-occurs_with	与……同时发生
course	process_of	……的进程
using	uses	使用
has_definitional_manifestation	manifestation_of	具有明确的表现,表现为
method	method_of	……的方法
severity	degree_of	……的程度-等级
has_focus	issue_in	……之中的焦点问题

词是指表示术语概念的属性,包括概念的历史属性、概念的模型属性、以及未经批准的属性等;背景依赖性分类词主要包括的是临床电子病历里所涉及的背景性的词,比如:家族史、过敏史、治疗反馈等。

从4个语言系统的历史发展情况来看,每个语言系统的语义关系均是在采纳同时期其它系统语义关系的基础上,根据自身研发需要添加新的语义关系,从而形成符合各自需要的语义系统。

四、中医药领域语言系统及语义关系发展现状分析

1. 中医药领域语言系统发展现状分析

(1) 同一概念的术语表达形式多样。

中医药领域中一词多义、多词同义现象非常普遍,这种术语表达形式多样化的现状使术语规范工作显得尤为重要,中国中医科学院中医药信息研究所基于这一形势设计了TCMLS,TCMLS依据UMLS的设计理念较好地解决了中医药领域多个相近术语概念之间的混乱问题。

(2) 不能满足电子病历的发展。

电子病历的研制和开发需要一部面向临床的专业较强的术语集来支撑,它既可以满足用户智能化录入临床数据的需要,又能够用于优化自然语言处理,还能完成医学信息的检索、提取与分析,而TCMLS由于服务方向和开发的侧重点有不同,难以满足电子病历发展的需要。为现代医学电子病历提供临床参考术语的SNOMED的发展为中医药临床术语的建立提供了新思路,中国中医科学院信息所于2006年在借鉴UMLS、TCMLS和SNOMED的研究基础上,意在开发一部具有中医特色,适合中医临床电子病历发展的术语集,为电子病历的发展提供支撑。

2. 中医药领域语义关系存在问题

(1) 沿用西医语言系统语义关系。

从中医药语言系统的总体发展来看,TCMLS和中医临床术语集均沿用了所对应西医语言系统的框架结构和语义关系。

虽然中医临床术语集、TCMLS、SNOMED和UMLS同属医学领域,有其相近的学科属性,但中西医有着不同的理论体系和实践,中医有其特有的医疗过程,不可避免地会涉及到一些行业领域内所特有的问题。虽然TCMLS新增4个中医语义关系,但中医临床术语集主要是为中医临床信息服务的,与TCMLS的服务方向有所不同,因此TCMLS所增加的语义关系在

表达临床层面的语义关系时有其自身的局限性。

不仅如此,西医医学语言系统的语义关系的发展虽然已日趋成熟和完善,但每个语义关系的自身作用和使用范围并不明确,虽然各个系统已给出了每个语义关系的定义,但并未对具体的应用范围做出明确规定,因此每类关系所应该承载的“量”并不清晰,对每一级语义关系所能连接的语义类型范围亦不明确,导致每个关系不能清晰地各司其职地连接相应的语义类型,尤其是在UMLS中三级和四级语义关系“度”的把握上比较模糊。

(2) 缺乏应用和实践。

两套中医药语言系统(TCMLS和中医临床术语集)还只停留在术语研究的层面上,没有与实际应用相联系,因此中医临床术语集并未按其特有的服务目的,对即将采用的语义关系在真实的病历中考察过其适用性,还没有经过最终实践的检验和认证,需要进一步在应用中改进与完善。

3. 建议

在中医临床术语集语义关系的建设上,建议收集大量中医医院病历、中医期刊论文收录的病历或中医典籍中的医案,分析病案中凸显临床特色医疗关系的临床术语,并提取和分析概念间具有临床特色的语义关系。

参考文献

- 1 张金凤,张燕飞,陈玉顺.受控语言与自然语言结合模式比较研究.图书情报知识,2005,4:75-77.
- 2 刘红阁,郑丽萍,张少方.本体论的研究和应用现状.信息技术快报,2005,3(1):1-12.
- 3 刘炜,李大玲,夏翠娟.基于本体的元数据应用.图书馆杂志,2004,23(6):75-77.
- 4 方青.基于本体论的中医药一体化语言系统.浙江:浙江大学计算机科学与技术学院计算机应用系,2004:3.
- 5 贾李蓉.中医药学语言系统语义关系初探.北京:中国中医科学院,2005.
- 6 杨继先.电子病历特征及其设计方法探讨.中国医院管理,2003,23(5):44-45.
- 7 SNOMED CT [OL]. <http://www.snomed.org/snomedct/index.html>, 2007-3-27.
- 8 尹皎,郑军. Ontology 在语义信息检索中的使用. 农业图书情报学刊, 2006, 18(12):119-121.
- 9 SNOMED Clinical Terms® (SNOMED CT®) Report-2 [OL]. http://gpa.senet.com.au/content/res_content/current/vocab/2_02_2.pdf, 2007-3-27.
- 10 Fact Sheet Unified Medical Language System® [OL]. <http://www.nlm.nih.gov/pubs/factsheets/umls.html>, 2007-3-27.

The function and status of semantic relationships in language systems

Yang Yang, Cui Meng, Li Yuanbai

(Institute of Information on Traditional Chinese Medicine, China Academy of
Chinese Medical Sciences, Beijing 100700, China)

Abstract: This paper reviews the development of the semantic relationships in the four main language systems, and discusses the issues existing in the semantic relationships of the medical language system. It proposes to extract semantic relationships from the clinical medical records, which may better express the special relationships in clinics.

Keywords: Traditional Chinese Medicine; Medical language system; Semantic relationship; overview

(责任编辑:李沙沙,责任译审:张立巍)

医药卫生体制改革

国务院深化医药卫生体制改革领导小组会议强调, 扎实有序建立基本药物制度, 稳步推进公立医院改革试点。日前, 国务院副总理李克强主持召开医改领导小组第四次会议, 审议了《关于 2009 年实施国家基本药物制度工作方案》和《关于公立医院改革试点的指导意见》, 研究建立国家基本药物制度、推进公立医院改革试点等医改工作。会议强调, 基本药物制度和公立医院改革是医改的重点和难点, 涉及多方面利益关系的调整, 要周密安排, 精心组织, 把改革任务落到实处, 使人民群众得到实惠。在建立国家基本药物制度方面, 每个省(自治区、直辖市)将于今年年底前在 30% 的政府举办城市社区卫生服务机构和县基层医疗卫生机构实施公开招标采购、统一配送, 全部配备使用基本药物并实行零差率销售; 三年内, 政府举办的基层医疗卫生机构全部配备和使用基本药物, 其他各类医疗卫生机构按照规定优先使用基本药物。在推进公立医院改革试点方面, 今年各省(自治区、直辖市)将选择若干城市开展公立医院改革试点, 国家将在东、中、西地区各选择一批有代表性的城市开展试点。今后三年, 要总结试点工作经验、形成改革总体思路 and 主要政策措施, 在全国推进公立医院改革。

两部门联合发文进一步加强基本医疗保险基金监管。为进一步加强对基本医疗保险基金管理, 提高基金使用效率, 日前, 人力资源和社会保障部和财政部联合下发了《关于进一步加强基本医疗保险基金管理的指导意见》。《意见》要求, 要增强医保基金的共济和保障能力。加大医保扩面和基金征缴力度。各地要按照 3 年内基本医保参保率达到 90% 以上的目标, 进一步加大扩面力度。切实妥善解决关破国有企业退休人员参

加城镇职工基本医疗保险问题, 同时, 统筹解决包括关破集体企业退休人员和困难企业职工等在内的其他各类城镇人员的医疗保障问题。进一步加大医保基金的征缴和稽核力度, 确保基金应收尽收。逐步提高基本医疗保险保障水平, 减轻参保人员的个人负担。在精心测算的基础上, 适当提高政策内住院医疗费用报销水平, 逐步提高统筹基金最高支付限额, 规范门诊大病管理。鼓励开展居民医保门诊统筹, 有条件的统筹地区可探索调整城镇职工基本医保个人账户使用办法。提高基本医疗保险统筹层次, 加快推进工作, 到 2011 年基本实现市(地)级统筹。

三部门印发开展城镇居民基本医保门诊统筹指导意见。人力资源和社会保障部、财政部、卫生部日前联合发出《关于开展城镇居民基本医疗保险门诊统筹的指导意见》, 要求根据城镇居民基本医疗保险基金支付能力, 在重点保障参保居民住院和门诊大病医疗支出的基础上, 逐步将门诊小病医疗费用纳入基金支付范围。《意见》指出, 开展城镇居民基本医疗保险门诊统筹, 要在坚持基本医疗保险政策规定的基础上, 充分考虑门诊医疗服务特点和城镇居民对门诊医疗保障的迫切需要, 进一步完善基本医疗保险的保障范围、筹资、支付等政策和就医、费用结算、业务经办等管理措施, 通过统筹共济的方式合理分担参保居民门诊医疗费用。《意见》要求, 城镇居民基本医疗保险基金要坚持收支平衡的原则, 门诊统筹所需费用在城镇居民基本医疗保险基金中列支, 单独列账。开展门诊统筹应充分利用社区卫生服务中心(站)等基层医疗卫生机构和中医药服务。将符合条件的基层医疗卫生机构纳入基本医疗保险定点范围。

(文 摘)