

# 领域本体在区域医疗信息共享系统中的应用<sup>①</sup>

李小智, 罗铁清

(湖南中医药大学 信息科学与工程系, 长沙 410208)

**摘要:** 随着医疗领域信息化建设不断推进, 为了提高医疗质量, 合理有效利用医疗资源, 医疗信息共享的要求日益强烈. 语义异构是医疗信息资源共享面临的一个主要问题, 提出利用领域本体整合分布式异构数据库资源, 介绍了医疗本体的构建和应用系统的设计, 为实现区域医疗信息共享提供了一种有效途径.

**关键词:** 医疗信息; 语义异构; 领域本体; 共享系统

## Application of Domain Ontology in Regional Medical Information Sharing System

LI Xiao-Zhi, LUO Tie-Qing

(Information Science and Engineering, Hunan University of Chinese Medicine, Changsha 410208, China)

**Abstract:** Following the informatization construction continues to advance in the medical field, in order to improve health care quality, and to use medical resources effectively, medical information sharing requirements become increasingly strong. Semantic heterogeneity is a major problem facing medical information resources sharing, this paper proposes the use of domain ontology integration of distributed heterogeneous database resources, describes the construction of medical domain ontology and design of application system for the realization of regional health information sharing provides an effective way.

**Key words:** medical information; semantic heterogeneity; domain ontology; sharing system

### 1 前言

近年来, 许多医院投入大量人力物力开展医院信息化工作, 根据自身的业务流程特点通过多种形式开发出了不同的医疗信息系统. 近年来在医院信息化建设中我国还没有完善的标准, 有许多标准并没有完全被执行, 各医院的信息化建设处于一种各自为战的状态, 医院内各系统间以及各医院间医疗数据不能共享, 导致出现以下问题: ①病人的检查数据和诊断结果不能跨医院共享, 在不同医院就诊需要重复检查, 增加了病人负担; ②在不同医院就诊因医生得不到连续完整的患者病历记录, 容易造成重复开药、过度医治, 医疗质量得不到保证; ③卫生行政管理部门无法对医疗行为进行实时监控, 大量积累在医疗卫生机构的有价值的信息无法实现增值. 因此, 区域医疗信息系统的互联互通和数据共享, 对于降低就医成本、提高医疗

质量、制定医疗政策有着重大意义.

医院信息系统现正从单一功能向多功能系统集成方向发展, 处理和储存数据规模越来越大, 这些数据不仅包括病人基本信息, 还包括大量的检查、化验、诊断、治疗方案等医疗信息. 由于医院之间使用的管理软件和数据库不同, 甚至同一家医院不同科室使用的软件可能也不同, 直接导致了医疗数据结构、存储方式的差异, 其主要体现在三个方面: ①结构异构, 主要指医疗数据有不同的结构形式, 既有保存病人、医护人员、药品信息的结构化关系型数据库数据, 又有以 XML 形式保存检查化验结果的半结构化数据, 还有如自由文本形式的病历记录、B 超、透视、电子显微镜影像资料的非结构化数据; ②语法异构, 因不同医院信息系统可能由不同软件公司开发, 数据存储和读取使用了不同的编程语言; ③语义异构, 因各医

① 基金项目:湖南省教育厅课题(11C0944);湖南中医药大学校级青年课题(99820001-70)

收稿时间:2014-04-22;收到修改稿时间:2014-05-30

院业务要求不同和医务人员专业素养有差异,使得同一个医疗术语概念在不同数据源中的表示各不相同,如“获得性免疫缺陷综合征”,其不同的名称就有“获得性免疫缺损综合征”、“艾滋病”、“爱滋病”、“AIDS”。

## 2 构建医疗领域本体整合异构数据资源

在医疗领域中,医院信息系统软件开发商希望有一种编码系统来满足临床电子病历发展的需要,行政管理部门需要一套标准的术语集用以理解和评价不同医院诊断/临床记录,医保部门也需要标准的术语编码实现对病人临床医疗数据进行自动化处理<sup>[1]</sup>,医疗数据的标准化、规范化是医疗行业信息化发展的必然要求。从技术层面来说,不同医院之间的医疗数据的异构性是制约区域内医疗信息共享主要原因,目前的技术如元数据可以有效解决结构异构和语法异构,但无法解决语义异构,解决不同数据源的语义异构是实现医疗信息共享面临的一个主要问题。

医疗领域本体是对医疗领域知识进行规范化的抽象与描述,不仅可以清晰的表达领域内有关的基本概念,还能够揭示基本概念之间丰富的关系,表达复杂的语义,能够促进该领域内不同主体之间的语义交流<sup>[2]</sup>。在区域医疗信息共享系统中运用领域本体,确定规范的概念描述和统一的医学术语表,可有效解决不同数据源的语义异构问题,并且通过医疗本体进行语义分析,提高对医疗信息搜索、整合与共享的效率。

### 2.1 医疗领域本体的构建

构建医疗领域本体要完成以下工作,依据医疗领域知识体系建立医学概念类,定义概念类之间的关系,概念类之间的关系包括类的医学属性、属性值及其实例,获取医学公理并用这些公理对医学知识进行一致性分析和知识推理。医疗领域本体中的概念是医疗领域公认的、规范化的描述,可以从领域专家、非结构化文本、结构化的数据库数据中获取。本体中的概念类之间有四种基本的关系,part-of:表达概念之间部分与整体的关系;kind-of:表达概念之间的继承关系;instance-of:表达概念的实例与概念之间的关系;attribute-of:表达某个概念是另一个概念的属性。如图1为医嘱对应的本体示例。

医疗本体的构建过程中,可选择Protégé为建模工具,OWL为本体描述语言,采用斯坦福大学医学院的七步法为主要参考方法来完成医疗本体的具体技术实

现。由于医疗领域的概念繁多、关系复杂,建立医疗领域本体是一项非常复杂的工作,要求本体设计人员要有深厚的医疗领域相关背景知识,同时又要求有很扎实的计算机专业知识。为确保概念和关系描述的充分性、规范性、准确性,医疗领域本体的设计和检验需要医疗领域专家参与,也需要在使用过程中不断加以完善和维护,因此医疗领域本体的构建是一个持续、反复的过程<sup>[3]</sup>。

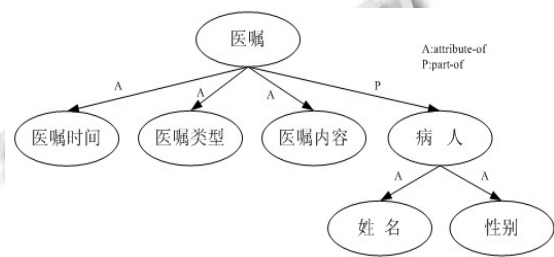


图1 医嘱对应的本体示例

### 2.2 医疗领域本体与数据库映射

区域医疗信息集成共享针对医疗领域数据共享并且数据源较为稳定,采用全局领域本体。全局领域本体对应于各分布、异构数据源,作为所有异构数据源的通用语义模型,即所有的异构数据源都直接映射到全局本体上。

当前大多数医院信息管理系统中的数据存取采用关系型数据库为主,非结构化数据如图片、影像大多是以保存存储路径方式保存在数据库中。医疗信息集成共享采用本体—数据库映射机制,从关系数据库的逻辑模型中自动抽取本体<sup>[4]</sup>,得到本体之后利用映射机制手动或半自动生成本体数据库映射表,映射表的建立需要充分理解关系库的设计和本体的语义,所有这一切均需要领域专家的参与。从关系数据库到本体映射方式有多种,其中简单映射是把本体类跟数据表对应,类的属性跟数据表中的列对应<sup>[5]</sup>。例如,医嘱本体与不同数据库的数据表映射模型如图2。

图2中展示两个不同数据库中所保存的医嘱信息数据表映射到医嘱本体类,记录医嘱时间、类型、内容的列分别映射到对应的医嘱本体类各属性,关系数据库中医嘱数据表通过外键与病人信息数据表联系起来,本体中采用概念之间的关系表示。这样通过映射机制,不同数据库虽然采用了不同的表名和字段名,但都指向了本体中的相同的概念。

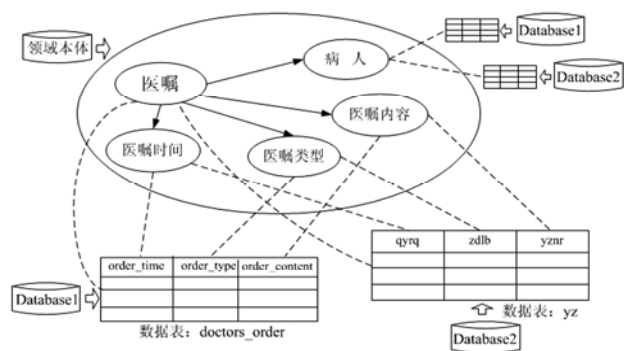


图2 领域本体—数据库映射模型

本体—数据库映射表详细记录了领域本体中的各个节点在不同数据库中与之对应是表名还是字段名以及在每个数据库中采用的名称,如表1所示。实际应用中,可以用XML文件保存映射表,也可保存在数据库中。如果数据库或数据库表结构发生变化时都需要相应修改映射表,维护更新可以通过手动来完成。

表1 本体—数据库映射表

本体概念	节点类型	Database1	Database2
医嘱	表名	doctors_order	yz
医嘱时间	字段名	order_time	qyrq
医嘱类型	字段名	order_type	zdlb
医嘱内容	字段名	order_content	yznr
病人	表名	patient	br

### 3 领域本体区域医疗信息共享系统的设计

区域医疗信息共享系统建设目标是搭建一个区域级医疗信息共享和交换平台,区域内的医疗卫生机构是该平台数据源的主体,医疗数据共享是该平台服务的核心。系统采用中介器(mediator)模式<sup>[6]</sup>,信息共享平台利用领域本体概念来统一描述用户的查询请求,即用户只需提供一个针对本体的查询请求,领域本体作为一个代理将用户查询请求转换成相应数据源的查询语句,从各个数据源中得到的结果返回给共享平台,并集成这些结果数据,再构造出用户所需的数据模式,以此实现分布式异构数据源的数据整合,底层异构的数据源对用户来说是透明的,用户是不知道数据来自于哪个数据库和数据源格式的。各异构数据源采用分散自治模式,医疗信息数据分布存储在各医疗机构端,信息共享平台不存储实际数据,只为平台用户如医院、社区卫生站、医疗行政管理部门提供医疗信息数据的查询服务,是逻辑视图上统一但是在物理上分散

存储的一种设计,基于领域本体的区域医疗数据共享系统的体系结构如图3。

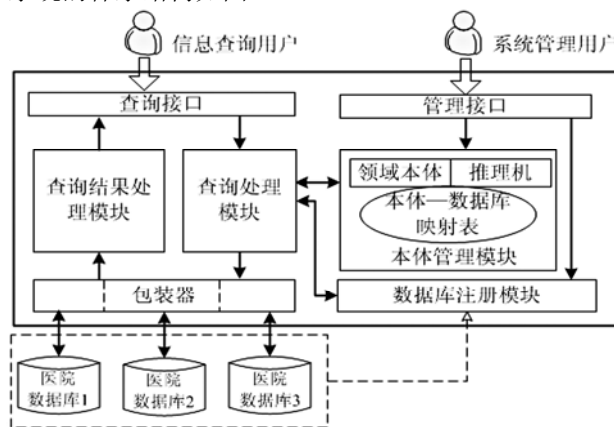


图3 区域医疗数据共享系统的体系结构

该系统主要分为三个层次,分别是数据层、业务逻辑层、表现层。数据层包括各医疗机构的异构数据源,如Oracle、Sql Server、DB2,业务逻辑层提供了数据共享集成业务的逻辑服务,它是系统核心层,负责接受查询请求、查询语句的构造与分发、查询结果的集成、领域本体的管理等,表现层用来与用户进行交互的用户接口。

图3中主要组件功能如下:

(1)本体管理模块:完成领域本体的管理,如对医学本体类和属性定义和修改、医学公理的表述和更新,对本体与数据库映射表的的更新和维护。

(2)数据库注册模块:保存的各异构数据库类型、账号、IP地址等信息,这些信息由各医疗机构信息系统管理员提供,由共享平台管理员进行录入和维护。

(3)查询处理模块:接受查询请求,根据推理机的语义分析结果检索本体—数据库的映射表,找到所要查询的数据是在哪个数据库的哪些数据表中的哪个字段,并根据数据库注册信息将查询转换成不同数据库的逻辑查询语句。

(4)包装器:异构数据库具有各自不同的包装器,包装器具有查询语句转换功能,将逻辑查询语句转换成基于不同数据库的查询sql语句,并发送给相应的数据库执行;查询结果需要包装器进行转换才能集成处理,如数据类型的转换,使得不同数据库的数据可以进行运算。

(5)查询结果处理模块:将查询返回的结果集进行合并过滤处理,提供一个统一格式的数据返回给查询

接口。

(6)查询接口:为用户提供查询信息的操作界面和查询结果的显示界面。

(7)管理接口:为系统管理用户提供本体编辑、保存、备份还原以及数据库注册信息管理的用户界面统称。

数据查询流程如下:

用户提出查询请求,查询处理模块将用户提交的原始查询转换成统一标准的本体查询,推理机装载全局领域本体对查询的概念进行语义推理,匹配相近或相等的概念,再根据推理机的结果检索本体—数据库的映射表,找到所要查询的数据所属数据库信息,将查询转换成逻辑查询语句,然后通过包装器将逻辑查询语句转换成各异构数据库的查询 sql 语句并发送给相应的数据库执行。各包装器将底层各数据库查询的结果进行转换,再返回给查询结果处理模块,将查询返回的结果集进行合并过滤处理并返回给查询结果显示界面。

由于该系统是基于本体的共享平台,本体的推理机制使得系统查询功能具有语义处理能力。系统将用户输入的关键词转换为系统能理解的领域本体中概念,然后经过本体的推理,对用户的查询请求进行语义分析和语义扩展<sup>[7]</sup>。其中非常典型的推理应用就是,使用概念和属性的继承关系对查询进行扩展,如本体中定义,冠心病继承心脏病,对心脏病的查询可以扩展到对冠心病的查询,以提高医疗数据的准确率和查全率。

## 4 总结

随着医疗信息系统的发展,为了解决信息孤岛,充分利用各异构系统中的医疗信息,运用本体技术实现异构数据库数据共享,它可解决信息集成中的语义异构问题。本文提出了一种基于领域本体医疗信息共享系统设计方案,原型系统的实现是下一步工作的重点。该系统实际应用过程中,还有很多有待解决的问题,如构建医疗领域本体是一项非常繁杂的工程,为减少工作量,如何从异构数据源实现本体的自动抽取、本体—数据库动态实时映射等,这些问题也是下一步研究的重要课题。

## 参考文献

- 1 陈焱,姜慧敏.医学领域本体研究现状.情报科学,2006,24(10):1587-1590.
- 2 许刚.基于领域本体的基础地质异构数据语义共享研究[硕士学位论文].南京:南京大学,2011.
- 3 徐彬锋,温志浩.基于本体的医学知识库构建及应用.北京生物医学工程,2011,30(6):618-623.
- 4 李庆忠,王栋.基于语义网格环境中异构数据资源整合的研究.南京大学学报,2006,42(2):141-147.
- 5 唐富年,姚莉,李金洋.基于本体的关系数据库访问研究进展.小型微型计算机系统,2011,32(3):390-395.
- 6 刘桂斌.异构数据库的数据共享集成中间件技术研究[硕士学位论文].西安:西北工业大学,2005.
- 7 赵修文.基于本体的医疗搜索引擎的设计和实现[硕士学位论文].长沙:国防科技大学,2008.