



近两年国外本体应用研究进展

Review of Foreign Applied Research on Ontology from 2006 to 2008

徐 静 (北京大学信息管理系 北京 100871)

孙 坦 (中国科学院国家科学图书馆 北京 100091)

黄飞燕 (中国科学院国家科学图书馆 中国科学院研究生院 北京 100091)

[摘 要] 近两年国外本体应用研究主要集中于本体在信息检索、信息抽取、异构信息系统的互操作和集成、语义 Web 4 个领域的应用,并表现出以下特点:本体在信息检索、信息抽取、异构信息系统的互操作和集成 3 个领域的应用研究以方法论探索为中心,研究方法以算法、建模和系统架构为主;在语义 Web 的应用研究中,以用户为中心的开发思想逐渐得到重视;研究成果多为学术会议论文,期刊论文较少。希望通过对近两年国外本体应用研究进展的梳理,能够对国内相关研究有所启示。

[关键词] 本体 信息检索 信息抽取 异构系统

[中图分类号] G250; TP391 [文献标识码] A

[Abstract] Over the past two years, the foreign applied research on ontology mainly focused on the application of the ontology in the fields of information retrieval, information extraction, interoperability and integration of heterogeneous system, and Semantic Web. The study shows characteristics as follows: firstly, methodology exploration is a key in the study of the first 3 fields, meanwhile algorithms, modeling and structure-based systems are the principal methods; secondly, customer-centered research get more and more attention in the application study of ontology in Semantic Web; and thirdly, academic conference papers serve as the primary form, while journal papers are relatively less. We'd better get some reference for the domestic research from the foreign applied research on ontology.

[Key words] Ontology; Information retrieval; Information extraction; Heterogeneous system

1 引 言

随着互联网应用逐渐向智能化和自动化方向发展,本体作为概念模型和概念间关系的规范描述,被广泛应用于信息系统和下一代互联网语义 Web 领域。本体在信息系统领域的应用,主要集中在信息检索、信息抽取、异构信息系统的互操作和集成,为其提供语义层的操作支持;另外,本体作为一种能在知识层提供知识共享、重用的工具,在语义 Web 中的应用研究也较为普遍。本文以 2006 年至 2007 年国外本体应用研究作为分析对象,概括近两年来本体在信息检索、信息抽取、异构信息系统的互操作和集成、语义 Web 4 个领域的应用研究进展和特点。

2 近两年国外本体应用研究概况

笔者利用 ACM、LISA、ProQuest、Springer 4 个数据库,

采用“Title = (ontology) AND (information retrieval OR seeking OR query) ”AND “ Time=(2006-2007) ”、“ Title = (ontology) AND (information extraction) ” AND “ Time=(2006-2007) ”、“ Title = (ontology) AND (heterogeneity OR integration) ”AND “ Time=(2006-2007) ”、“ Title = (ontology) AND (semantic web) ” AND “ Time=(2006-2007) ”这几个检索式进行检索,加上通过网络检索和其他途径所获得的,经过去重和删除不相关文献,得到相关论文 98 篇。通过对这些文献的分析,试图描述近两年来国外学界在本体应用研究领域的现状。

3 本体在信息检索领域的应用研究

目前,主要的信息检索方法是直接基于关键词,这种信息检索技术已不能满足用户在语义上和知识上的需求,于是,寻求新的方法也就成为目前研究的热点。本体具有良好的概念层次结构和对逻辑推理的支持,因而在信息检

索,特别是在基于知识的检索中得到了广泛的应用^[1]。

3.1 单领域本体在信息检索中的应用

目前,国外对本体在信息检索领域的应用的研究较多,不少研究利用单个领域本体来改善特定学科领域的信息检索效果。休斯顿大学的 Ping Chen 和 Rakesh Verma 提出一种基于用户询问的文本摘要技术来改善医学信息检索效果。他们选择美国国家医学图书馆开发的 UMLS(Unified Medical Language System, 一体化医学语言系统)作为其主要的医学本体知识库, UMLS 能帮助医学信息系统理解生物学和健康领域的概念和术语的内涵以及相互之间的关系。首先,研究人员选取数个在线医学数据库作为医学资源信息来源,并通过网络爬虫从所选取的数据库中抽取文档并保存在本地文档库中,利用 UMLS 对本地文档进行数据清洗,生成新的文档摘要,建立基于概念的文档索引;其次,利用 UMLS 对用户输入的关键词进行关键词扩展,增加缩写关键词、语义相近或相关的医学术语等,删除冗余关键词,并反馈给用户,由用户最后决定;最后,根据用户最终确定的扩展关键词对本地文档库的文档摘要进行语义匹配度运算,检索结果按相关度高低排列返回给用户^[2]。

3.2 多领域本体在信息检索中的应用

此外,也有研究采用多本体整合的方式提高信息检索性能。Aijuan Dong^[3]提出一个基于多本体的多媒体注解模型,通过对领域无关的多媒体本体与领域本体进行整合,实现从多个特定领域角度对多媒体内容进行描述,从而使用户能更好地表达其对不同多媒体信息的需求。研究者基于多媒体内容描述工具 MPEG-7(Moving Picture Experts Group, 活动图象专家组)开发一个多媒体本体,提出一个整合多领域本体的策略,设计一个术语抽取程序来自动从多媒体数据的文本资源中抽取特定领域的本体术语。初步实验表明,基于多本体的多媒体注解可以更好地满足不同用户的信息需求^[3]。

高质量的本体知识是提高信息检索和管理质量的关键因素。多数研究都是选择现成的领域本体作为其本体知识库。本体在信息检索中的应用主要在以下两个环节:一是根据本体对被检索的信息资源进行文档预处理,包括语义标注与建立基于概念的文档索引;二是系统借助本体知识库分析用户查询中所包含的术语含义,准确理解用户的查询,从而提高信息检索系统的查全率和查准率。在用户检索过程中,系统可为用户提供语义提示,如进行关键词扩展、去除冗余关键词,并反馈给用户由其选择,使用户更精准地描述自身的息需求,与用户建立良好的互动关系。

4 本体在信息抽取领域的应用研究

基于本体的信息抽取技术是一种把本体和信息处理技术结合起来实现信息抽取的技术,它是现在信息抽取研究的一个热点,已经有了很多研究。传统的信息抽取虽然能

抽取出实体,但是缺乏领域知识来识别抽取实体之间的关系。因此,在信息抽取任务中引入相应的领域知识——领域本体来指导抽取过程,将能有效地提高信息抽取的性能^[4]。

4.1 知识工程方法

国外近两年在自由文本信息抽取研究中,主要的研究方法可分为以下两种:知识工程方法和机器学习方法。知识工程方法是指由专家对本体进行分析、调整而人工制定规则、模板。美国德雷塞尔大学 Zhou Xiaohua 等研究者开发了一个名为 MedIE(MEDical Information Extraction, 医学信息抽取)的系统,目的是从大规模自由文本临床记录中抽取和挖掘出大量患乳腺疾病的病人信息。信息抽取工作共分为 3 步:(1)采用基于本体的方法进行医学术语抽取,使用言语模式生成候选术语,并确认是否在本体库中存在,为了弥补本体库的不完整性,人工补充其他语义相似的术语;(2)最主要的环节,进行关系抽取,人工建立抽取规则,利用连接语法解析器(link-grammar parser)对句子进行解析,解析结果以图表形式显示,有着较高精确率;(3)文本分类,采用基于自然语言处理的特征抽取技术与 ID3 决策树(ID3 决策树算法是数据挖掘中常用的一科方法)相结合的方法。虽然该研究在试验中取得较高的准确率和召回率,但它的实验样本规模较小,只有 125 条临床记录。当系统面临处理大规模且又多样化书写模式的临床病例记录时,它的信息抽取性能很可能降低^[5]。Chyan Yang 等开发了 ITSIES 系统(IT Product Specification Information Extraction System, 信息技术产品规格抽取系统),融合了自然语言处理技术(Natural Language Process, 简称 NLP)和本体概念。研究者首先随机选取有关 IBM 和 HP 数个 IT 产品的 34 个网页作为文档库;使用 ANNIE(A Nearly New Information Extraction System, 基于规则方法的英文信息抽取系统)对网页文档进行分词,此外,ANNIE 的默认功能可以识别出有关如地址、日期等的信息;在 JAPE(Java Annotations Pattern Engine, Java 标注模式引擎)中人工建立语法规则,对信息实体进行标注;基于 IT 产品特征对概念及其关系进行描述,最终构建成 IT 产品规格本体。ITSIES 系统在对其它品牌的 IT 产品规格信息抽取试验中表现良好^[6]。

4.2 机器学习方法

机器学习方法是指给出根据本体中的概念进行标注的例子文档集,通过机器学习的方法来推导模板和模板的自动填充知识库的规则。生物医学领域研究发展速度快,学术研究成果丰富,知识更新频率高,所以经常会发现以往一些不相关的主题产生了新的关系。鉴于生物医学这一现状,依靠人工来对本体概念及其之间关系进行一一标注是不切实际的,因而采用自动查询技术从大规模文本数据库中对相关文档进行信息抽取是非常必要的。Hua Xiaohua 等开发一个 SPIE 系统(Scalable and Portable Information Extraction System, 可扩展便携式信息抽取系统)。与以往使用

注解语料库进行训练不同,该研究只使用少量的种子样本。(1)SPIE 根据研究者提供的少量原始种子元组在生物医学数字图书馆中检索样本文档;(2)元组集将文档分成两种,包含至少一个元组的文档自动标识为“正文档”,一个元组都没有的文档表示为“负文档”;(3)将正负文档转换成一个适当训练的样本库,在训练样本上运行数据挖掘运算法则生成规则集,再将规则转换成一系列的查询,目的是检索新的有用文档;(4)利用本体实体和功能型标签从文档中抽取包含一对实体的句子,并从句子中寻找元组;(5)从元组中学习抽取模式,再利用学习到的模式识别更多隶属这一关系的新元组,研究者对生成的模式和元组进行检验,虚假模式将被删除,高度置信度的元组被留下作为下一轮的种子样本;(6)利用学习到的种子样本返回到第1步,到生物医学数字图书馆检索出一系列文档,整个程序一直循环直到没有新的元组被添加到关系库中。SPIE 系统能在大规模生物医学文档库中抽取许多二元关系,如蛋白质间交互作用、蛋白质与 DNA 交互作用等^[7]。

知识工程方法主要依靠人工编制规则来建立领域本体知识库,需要学科领域专家的参与,开发过程较慢,其优点是不需要对本体知识库进行经常性更改,抽取效率较高,但由于其对该领域以后出现新的概念或关系无法及时发现,一旦发生这种情况,信息抽取的召回率会降低。在国外,近几年机器学习方法在信息抽取领域的应用研究受到广泛的关注,它主要通过学习已经标记好的语料库获取规则,无需投入大量人力。这种方法比知识工程方法快,但需要足够数量的训练文档,才能保证其信息抽取的效率。

5 本体在异构信息系统互操作和集成领域的应用研究

分布式网络环境下,无论是数字图书馆等学术应用系统,还是企业的商用信息系统,一定程度上都面临着系统异构、语法异构、模式异构和语义异构4个层面的异构问题。本体作为概念模型的明确的规范说明,正在成为解决异构系统互操作和集成问题的基础解决方案。

近两年,国外学者关于应用本体解决信息系统异构问题的研究主要围绕方法层面:探寻基于本体的有效方法,解决特定领域的异构问题。所采用的方法主要为本体与代理结合的方法、本体间映射方法。

5.1 基于本体的代理方法

基于本体的代理方法提供了一套解决异构问题的较为完整的框架。Huong 使用模糊逻辑决策算法,为今后的异构无线网络构造了一个基于本体的情景管理代理,用于管理无缝移动通信系统间漫游服务的网络接入问题^[8]。Irawati 分析在三维多通道交互框架下,构造利用空间本体实现语义集成的代理架构。该架构定义了空间本体,讨论其使用策略,以协助用户在诸如网络游戏等虚拟环境中进行对象寻址。关键问题是多通道输入的语义集成。研究者提出一

个多通道交互框架实现(Multimodal Interaction Framework, 简称MIF)。通过该框架,用户的多通道的输入信息被整合,系统输出语义集成的用户意愿,同虚拟世界(如网络游戏)进行交互操作。MIF由5个组件系统实现:(1)输入识别组件,将多通道的输入编码作为系统处理的基础,能识别包括文本、键盘操作、鼠标操作、按钮、标签、过程跟踪器等各种通道的输入信息;(2)特定领域功能组件,包含对特定领域的输入的处理函数;(3)交互技术功能组件,支持如3D光标、闪光灯的交互选择和操作技术;(4)对象本体库,是领域概念、规则、关系的集合;(5)交互管理机器人,是核心组件,对前两个组件的处理结果,利用交互技术功能组件的交互选择和操作技术,与对象本体库关联,获得语义集成的交互式输出^[9]。

5.2 映射方法

对基于本体的领域知识和本地信息资源之间的互操作和集成问题,研究者主要通过本体间的映射方法来分析解决。Park 分析提出了电子商务环境下,异构的产品分类目录间的本体映射方法。由于严格的映射策略可能会导致客户的查询失败,文章提供了一个映射算法,确保在合理的损失精度下,增加查询结果中的产品组合。通过与本体映射算法 PROMPT 的比较实验,此算法的平均查询结果数高于 PROMPT 46.5%,平均精度比 PROMPT 差 13.6%,成功地实现了查询结果的显著提高与精确度的合理降低^[10]。基于本体的异构信息查询中,分解和重构一个本体概念实例的通用查询,以形成多个基于本地信息源的子查询,是一项基础性挑战。Li Jian^[11]开发了一套基于本体的异构信息集成的查询分工和重构方案,主要采用基于映射的查询分解和集成。其中,通用本体被用来描述用户感兴趣的领域知识,用户通过查询通用本体获得本体概念的实例。该方法分为3步:(1)提供一种统一的概念实例来表述查询。假设查询结果是一个元组集合,定义了一系列基于元组的操作,因此查询可以表述为对于元组集合的各种操作。(2)将通用查询分解为多个本地查询的方法,分解基于映射。此方法基于两个假设:首先,通用映射由两个本地映射组成,以集成两个本地信息资源;其次,由多个本地映射组成的通用映射的查询分解,可以通过前面假设的二元结构的多次过程迭代实现。基于此,查询将通过对应的映射实现,研究人员提供了一种将通用映射下的通用查询分解为本地映射下的多个子查询的算法。子查询对应于元组集合的各种操作,其操作结果的集合为对应的通用查询结果。(3)本地的查询结果被集成为通用查询的结果。集成的主要工作是组合分解了的元组的各种操作结果^[11]。

另外,软件工程和数学领域的相关方法也被广泛借鉴,其中本体与机器学习的整合、贝叶斯网络等方法的使用渐起。Dou 设计了一个架构,使用基于本体的模式表示、一阶

逻辑推理、SQL 套件集成了两个样本关系数据库,验证了表示和推理技术分析的有效性^[12]。Tsymbol通过重新界定概念相似性,整合服务于大型多维特定疾病的异构数据的本体提供的知识,提高机器学习的性能^[13]。McGarry通过文本挖掘 II 型糖尿病的近期研究文献,利用贝叶斯网络萃取出尚未关联到的新信息,结合基因本体和数据库内已有的知识,实现本体的自动化生成或演化^[14]。随着本体数量的增长,本体的管理问题得到研究者重视。利用本体进行异构核查、预测的研究也逐渐兴起,不再赘述。

在 4 类异构问题中,研究者普遍较为关注语义异构问题,其中基于本体的代理方法提出了解决语义异构和模式异构的较完整的架构,本体间映射的方法多侧重解决语义异构问题。语义异构问题的解决,使交互信息的含义得到了明确表达,交互双方能够对交互信息进行推理以实现更有效的语义互操作,推进 Web 服务演进为语义 Web 服务。研究方法上,主要采用软件工程和数学建模方法,已被开发出的特定领域本体(如基因本体)已经成为重要的研究对象。

6 本体在语义 Web 中的应用研究

6.1 语义 Web 中的本体开发研究

语义 Web 的提出源于如何为 Web 上的资源提供计算机可以理解的内容。本体作为共同标准的概念体系,支持简单推理,促进计算机相互理解和互操作,本体的应用可有效促进语义 Web 的性能。

近两年,研究者在语义 Web 中的本体开发研究集中于引入多种方法,提高对不确定和非精确信息的表示能力,促进半自动化、自动化本体生成和本体演化,以提供智能化的语义 Web 服务。Yu Qing^[15]、Tho Quan Thanh^[16]均通过引入模糊逻辑,提出了 FOWL(Fuzzy Web Ontology Language, 模糊网络本体语言)和 FOGA(Fuzzy Ontology Generation framework, 模糊本体生成框架),自动生成模糊本体以描述现实世界的非精确性或不确定性知识。现有的少量本体不足以满足海量信息和语义 Web 成长的需要, J. R. G. Pulido 采用 SOM(Self-Organizing Maps, 自组织特征映射)方法,从自动抓取的网页文档中做出知识地图;在知识地图中,识别出本体的组成部分,将其组织成实体、关系和功能 3 部分,服务于语义 Web 软件代理的动态智能操作^[17]。关于本体演化和本体再造可动态改进本体,研究者在其可行性和一致性方面给予了关注。Su-Kyoung Kim 论证在语义 Web 中,应用基于推理的网络本体后,在查全率和正确率上高于基于注解的本体应用^[18]; Kang Seung Hwan 论述了将信念修正理论整合进入本体再造方法,可确保本体修正的一致性^[19]。

6.2 语义 Web 中的本体功能系统开发研究

相关本体功能系统的开发研究,已经从本体映射和本体演化研究扩展到本体登记、本体清洗、本体映射、异构

整合、广告匹配和基于本体的评价等本体功能流程的多个方面,显示本体应用研究已具备一定的深度和广度。

本体在海量复杂信息资源组织表示上的便利性和可靠性,首先基于高效的本体登记。Wang PChong 分析提出了 MF14Onto(Metamodel Framework for Interoperability: Metamodel for Ontology Registration, 本体登记的互操作元模型框架)方案,提供了一个基于信息系统和本体之间的语义层互操作的共同框架,以登记本体及其演化信息^[20]。本体清洗可在知识库中发现并修正本体和数据源中的异常。Jose A 从清洗代理设计的角度出发,指出在面临知识库完整性、连续性和正确性的不确定情况下,只有基于可鉴定的推理过程才能获得逻辑可信的清洗^[21]。有效的本体映射可以更好地理解用户需求。Vanessa Lopez Garcia 分析指出 AquaLog 本体系统——一种本体驱动的门户式自动问答系统。不同于以往的扩展用户查询词的问答系统,该本体的处理过程为:用户输入自然语言查询词,采用关系相似性服务组件(Relation Similarity Service Component, 简称 RSSC)将自然语言查询词转换成本体作为系统输入,返回的答案来自一个或多个知识库,实现不同的答案策略的组合^[22]。Pa Pa Nyunt 提出了基于本体的代理人社区的做法,通过动态生成通用本体,为信息集成系统演示了一种灵活和动态的方法来融合异构数据源,以最大限度地进行信息共享^[23]。Jorge Cardoso 开发了一种匹配语义网络服务请求与语义网络服务广告的算法,其中,匹配算法基于以本体概念形式组织的输入输出的相似度^[24],拓展了本体应用的商业价值。Natenapa Sriharee 提出了一种基于本体的评价模型以评价服务质量。服务的质量评价过程为:首先,第三方组织以本体值的方式分发分值表;其次,通过服务的被检索情况评分,主要通过检索匹配算法算出基于消费者喜好的服务质量的排名信息^[25]。通过这种服务质量评价,消费者可以精确定位出需求服务:即大多数消费者喜好的服务。

6.3 以用户为中心的本体应用研究

随着语义 Web 研究的深入,以用户为中心的开发思想逐渐得到重视。语义 Web 用户分为两类:最终用户和研究型用户。

在如何获取最终用户的个性化需求方面, Jiang Xing 通过构建用户本体来学习和表达用户的个性化需求:基于统计,使用领域本体和激活扩散程序,学习基于语义的用户本体模型,以表达用户的兴趣。通过在语义搜索引擎中利用用户本体搜寻学术刊物的实验验证,比之于普通关键字搜索、领域本体搜索,用户本体和激活扩散理论在提供个性化语义服务方面具有更好的有效性^[26]。在如何支持最终用户进行格式自由、自然语言的 Web 输入,提供高质量的 Web 服务方面, Al-Muhammed 构建了本体识别软件服务代理,该代理以任务本体为基本组件,任务本体由领域本体(编码领域信息,如对象集、关系集的规则和实例,对象集、

关系集的值的操作)和处理流程本体(一次性编码的领域独立的通用处理流程)构成。此代理可以识别出与领域本体相关的数据库的信息请求限制,当无法满足此类限制时,使用处理流程本体,为用户提供相关或近似相关的解决方案^[27]。

对于研究型用户的个性化创造的支持,也是本体研究的一个热点。W3C(The World Wide Web Consortium, WEB标准英文写法)的语义Web最佳实践和发展工作小组已成立一个专责小组,以支持用户自定义数据类型。W3C标准本体语言OWL DL(OWL Description Logic,网络本体语言的描述逻辑)的两个判定扩展OWL-Eu(OWL with Unary Datatype Expressions,OWL扩展的一元数据类型描述语言)和OWL-E(Extending OWL with Expressive Datatype Expressions,OWL扩展的数据类型描述语言),也都提出支持以用户为中心的数据类型和用户定制的数据类型谓词。Jeff Z Pan基于FaCT DL(Fast Classification of Terminologies,描述逻辑快速分类推理机),为这两种语义网络本体语言提出一个灵活的推理结构,该结构可实现:(1)允许用户定义自己的数据类型和基于内置式谓词的数据类型谓词;(2)新的推理数据类型可以被添加到该架构而无需做概念层次的变化^[28]。Akhilesh K Sinha介绍了一种名为ATHENE的基于本体的可视化语义Web流程建模工具,它允许用户以图形化方法开发过程模型,并自动生成过程模型背后的本体^[29]。

6.4 语义Web特定应用领域的本体应用研究

语义Web可以实现特定领域信息的有序组织、存储和检索,而本体通过对共同认可的概念、概念与概念之间关系的精确定义,支持语义层次上信息共享和交换,改进了特定领域的Web服务。Jari Veijalainen设计了移动通信服务中基于本体的语义Web服务架构,以提供创新型移动服务^[30]。Alfio Ferrara构建了一个MX-onto(Music Context ontology,音乐情境的本体),该本体使用OWL进行音乐信息的情境描述,定义分类体系和分类规则,在音乐资源和相应的音乐类型之间捕捉音乐流派、界定多成员之间的关系,从而支持基于情境和基于近邻的音乐资源搜索^[31]。Wang Xia借用企业治理架构(Governance Enterprise Architecture,简称GEA)中的领域本体,分析公共行政服务领域的概念映射情况^[32]。Frederico Fonseca探讨了从地缘现实到推导性本体下的地理空间语义网络的发展方向^[33]。

医学领域对患者病历、同一病种的不同疗法等资料的共享要求高,对于标准化、智能化的语义网络需求强烈。Hawthorne通过对医务志愿者的调查采访,验证语义Web本体下的电子健康记录作为脊柱治疗方法具有高可行性,创建了名为Chiro的本体系统,供使用脊椎指压治疗法和脊椎对抗疗法的医生共享患者信息^[34]。Sajjad Hussain分析了语义网络框架下的本体驱动型临床实践指南(Clinical Practice Guidelines,简称CPG)的创建和执行,通过引入一个领域

本体、一个病人本体、一个决策规则和一个规则执行引擎,采用本体驱动的语义Web方式构建电子化的CPG和临床决策支持系统(Clinical Decision Support Systems,简称CDSS)。CDSS系统构建分为4步:(1)定义临床实践指南的本体,以本体化显示指引元素模型(Guideline Element Model,简称GEM)编码的临床实践指南。本体定义基于指引元素模型中的文档类型定义(Document Type Definition,简称DTD)。(2)开发CPG决策逻辑定义工具,定义CPG的语法规则。(3)开发CPG执行引擎,通过JENA推理系统执行CPG的逻辑规则。(4)实施自动化决策树生成模型,提供解决方案的推理过程跟踪,以协助医生理解所提出的建议。实际应用中,给定一个病人实例,CDSS系统即可给出相关临床建议^[35]。

总体来说,本体在语义Web中的应用研究已具有一定的广度和深度,从自动化、智能化本体构建,拓展到本体功能系统的完善,开发思想也开始强调用户的个性化需求,语义Web本体的应用领域研究进一步拓展。研究方法上,算法开发、模型构建占主流,少数研究采用了用户访谈法,量化分析方法应用不多,说明本体在语义Web中的应用研究尚未达到成熟阶段,在研究方法和研究内容上均有较大提升空间。

7 国外近两年本体应用研究的特点

7.1 计算机科学领域的研究居多

有关本体应用的研究学者来源较广,包括计算机科学、图书馆学情报学、医学、地理学、通信、经济管理等多个领域,其中,计算机科学领域的研究者居多。近两年来,图书馆学情报学对本体的应用研究较少。可能由于本体的开发和应用多涉及到系统、数据库、知识库等的技术操作,加之本体本身的研究尚未成熟,在图书馆学情报学领域的应用尚处于起步阶段。

7.2 本体应用的领域广

近两年,基于本体的一系列应用研究的对象领域较广,包括中小企业的系统互操作、三维地理空间本体和视觉本体的生成、音乐等多媒体资源的文本分类和检索、移动语义网络建设和异构网络接入的集成、电子商务的分类目录互操作、医学领域的本体构建、知识抽取与共享。由于医学信息复杂,对病人的过往病历、同病种的不同疗法的知识共享要求高,该领域的本体研究较多,在现有的基因本体的基础上,进行了较广泛的应用拓展研究。

7.3 研究方法以算法、建模和系统架构为主

在研究方法上,研究者们普遍采用算法改进、数学建模或者系统架构的方法,服务于某一特定领域的本体开发,或者提高检索、抽取等的处理效率,解决特定领域的系统异构问题。由于本体开发需要领域专家的参与,界定规范化的概念,已有少数学者进行专家访谈和调查,分析基于本体系统的可用性与效率。随着本体应用的普及,此类面

向用户的研究方法将逐渐增多。

7.4 学术会议的相关研究居多，期刊论文较少。

关于本体应用的研究论文，以 IEEE、ACM、IASTED 等学术会议论文居多，期刊论文较少。说明关于本体的应用研究尚处于初始阶段，各类国际会议较为关注本体应用的相关研究，关于本体的研讨会较多。

参考文献：

- [1] 邓志鸿, 唐世渭, 张 铭, 等. Ontology 研究综述[J]. 北京大学学报:自然科学版, 2002(5):730-738.
- [2] Chen Ping, Verma Ra. A Query-Based Medical Information Summarization System Using Ontology Knowledge[C]. Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems table of contents, 2006:37-42.
- [3] Dong Aijuan, Li Honglin. Multi-ontology Based Multimedia Annotation for Domain-specific Information Retrieval[C]. Proceedings of the IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing, 2006:158-165.
- [4] 陈 静, 朱巧明, 贲正仙. 基于 Ontology 的信息抽取研究综述[J]. 计算机技术与发展, 2007(10):84-86.
- [5] Zhou Xiaohua, Han H, Chankai I, et al. Approaches to Text Mining for Clinical Medical Records[C]. Proceedings of the 2006 ACM symposium on Applied computing table of contents, 2006:235-239
- [6] Yang Chyan, Chen Liang-Chu, Peng Chun-Yen. Developing and Evaluating an IT Specification Extraction System[J]. The Electronic Library, 2006(24):832-846.
- [7] Hua Xiaohua, Lin Y, Song I, et al. A Semi-Supervised Efficient Learning Approach to Extract Biological Relationships from Web-Based Biomedical Digital Library[J]. Web Intelligence & Agent Systems, 2006(4):327-339.
- [8] Nguyen N, Matsumoto M. Ontology-Based Context Management Agent for Vertical Handoff Use Fuzzy Logic Decision in Heterogeneous Network[C]. Proceedings of the First International Conference on Complex, Intelligent and Software Intensive Systems, 2006: 215-220.
- [9] Irawati S, Calderon D, Ko P. Spatial Ontology for Semantic Integration in 3D Multimodal Interaction Framework[C]. Proceedings of the 2006 ACM international conference on Virtual Reality Continuum and Its Applications[C], 2006:129-135.
- [10] Park S, Kim W. Ontology Mapping Between Heterogeneous Product Taxonomies in an Electronic Commerce Environment[J]. International Journal of Electronic Commerce, 2007(2): 69.
- [11] Li Jian, Jin Beihong. Query Division and Reformulation in Ontology-Based Heterogeneous Information Integration[C]. Proceedings of the 15th International Conference on Computing, 2006:186-196
- [12] Dou Dejing, Paea L. Ontology-Based Integration for Relational Databases[C]. Proceedings of the 2006 ACM Symposium on Applied computing, 2006:461-466.
- [13] Tsymbal A, Zillner S, Huber P. Feature Ontology for Improved Learning from Large-Dimensional Disease-Specific Heterogeneous Data[C]. Proceedings of the Twentieth IEEE International Symposium on Computer-Based Medical Systems, 2007:595- 600.
- [14] McGarry K, Garfield S, Wermter P. Auto-Extraction, Representation and Integration of a Diabetes Ontology Using Bayesian Networks [C]. Proceedings of the Twentieth IEEE International Symposium on Computer-Based Medical Systems, 2007:612-617.
- [15] Yu Qing, Wang Jinlin. Extending Ontology Language for Semantic Web[C]. Proceedings of the 2007 International Conference, 2007: 116-119.
- [16] Quan Thanh Tho, Siu Cheung Hui, Fong A, et al. Automatic Fuzzy Ontology Generation for Semantic Web[C]. IEEE Transactions on Knowledge and Data Engineering, 2006:842 - 856.
- [17] Pulido J, Herrera R., Arechiga M, et al. Identifying ontology components from digital archives for the semantic web[C]. Proceedings of the 2nd IASTED international conference on Advances in computer science and technology, 2006:7-12.
- [18] Kim S. Implementation of Web Ontology for Semantic Web Application[C]. Proceedings of the Sixth International Conference on Advanced Language Processing and Web Information Technology (ALPIT 2007), 2007:159-164.
- [19] Kang Seung Hwan, Lau S K. Ontology Revision on the Semantic Web: Integration of belief revision[C]. Proceedings of the 40th Annual Hawaii International Conference on System Sciences, 2007: 61.
- [20] Wang Chong, He Keqing, He Yangfan. MFI4Onto: Towards Ontology Registration on the Semantic Web[C]. Proceedings of the Sixth IEEE International Conference on Computer and Information Technology, 2006:40.
- [21] Alonso-Jimenez J, Borrego-Diaz J, Chavez-Gonzalez A, et al. Foundational Challenges in Automated Semantic Web Data and Ontology Cleaning[J]. IEEE Intelligent Systems, 2006(1):42-52.
- [22] Garcia L V, Motta P, Uren V. AquaLog: an ontology-driven question answering system to interface the semantic web[C]. Proceedings of the 2006 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, 2006:269-272.
- [23] Nyunt P, Thein N. Ontology-Based Agent Community for Information Integration System in Semantic Web[C]. Proceedings of the First Asia International Conference on Modelling & Simulation, 2007:106-111.
- [24] Cardoso J. Discovering Semantic Web Services with and without a Common Ontology Commitment[C]. Proceedings of the IEEE Services Computing Workshops, 2006:183-190.
- [25] Sriharee N. Semantic Web Services Discovery Using Ontology-Based Rating Model[C]. Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, 2006:608-616.
- [26] Jiang Xing, Tan P. Learning and inferencing in user ontology for personalized semantic web services. Automatic Fuzzy Ontology Generation for Semantic Web[C]. Proceedings of the 15th international conference on World Wide Web, 2006:1067-1068.
- [27] Al-Muhammed, Muhammed J. Ontology aware software service agents: Meeting ordinary user needs on the semantic Web[D]. Provo: Brigham Young University, 2007.
- [28] Pan Z J. A Flexible Ontology Reasoning Architecture for the

- Semantic Web[J]. IEEE Transactions on Knowledge and Data Engineering, 2007(2): 246 -260.
- [29] Sinha K A, Singh N. ATHENE- An Approach for Modeling Semantic Web Processes over User-friendly and Editable Ontology Models [C]. Proceedings of the Third International Conference on Next Generation Web Services Practices, 2007: 93-98.
- [30] Veijalainen J, Nikitin P, Tormala V. Ontology-based Semantic Web Service platform in Mobile Environments[C]. International Conference On Mobile Data Management Proceedings of the 7th International Conference, 2006: 83.
- [31] Ferrara A, Ludovico A L, Montanelli S, et al. A Semantic Web ontology for context-based classification and retrieval of music resources[J]. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP), 2006(3):177-198.
- [32] Wang Xia, Vitvar P, Peristeras P, et al. WSMO-PA: Formal Specification of Public Administration Service Model on Semantic Web Service Ontology[C]. Proceedings of the 40th Annual Hawaii International Conference on System Sciences, 2007: 96.
- [33] Fonseca F, Rodriguez A. Guest Editorial, From Geo-Pragmatics to

- Derivation Ontologies: New Directions for the GeoSpatial Semantic Web[J]. Transactions in GIS, 2007(3):313.
- [34] Hawthorne K E, Cohen S P. Chiro-informatics: towards a semantic web health record and ontology for coordination of patient information between chiropractic and allopathic physician[C]. Doctoral Thesis:Nova Southeastern University, 2006:216.
- [35] Hussain S, Abidi R. Ontology Driven CPG Authoring and Execution via a Semantic Web Framework[C]. Proceedings of the 40th Annual Hawaii International Conference on System Sciences, 2007:135.

[作者简介]

徐 静 女, 1986 年生, 北京大学信息管理系情报学硕士研究生, 发表论文数篇。

孙 坦 男, 1970 年生, 中国科学院国家科学图书馆副馆长, 研究馆员, 教授, 发表论文 60 余篇。

黄飞燕 女, 1984 年生, 中国科学院国家科学图书馆硕士研究生, 发表论文 7 篇。

[收稿日期: 2008-05-26]

(上接第 83 页)

- [4] Hiroki O, Kozo S. Method for Drawing Intersecting Clustered Graphs and Its Application to Web Ontology Language[C/OL]. Asia Pacific Symposium on Information. [2008-05-01]. http://www.jaist.ac.jp/~h_omote/research/apvis2006.pdf.
- [5] Rajugan R, Chang E, Dillon T. Visual Modeling of Ontology Views for e-Sciences Using XSemantic Nets [J]. Computer-Based Medical Systems, 2007. CBMS apos; 2007. Twentieth IEEE International Symposium, 2007(20-22): 601-606.
- [6] Hwang M, Kong H, Kim P. The Design of the Ontology Retrieval System on the Web[J]. Advanced Communication Technology, 2006. ICACT 2006. The 8th International Conference, 2006(20-22):1815-1818.
- [7] O'Brien P, Abidi S, et al. Modeling Intelligent Ontology Evolution Using Biological Evolutionary Processes[J]. Engineering of Intelligent Systems, 2006 IEEE International Conference, 2006(0):1-6.
- [8] Walid S, Saba. Ontology and Formal Semantics-Integration Overdue [EB/OL]. [2008-05-01]. <http://arxiv.org/abs/0712.1529>.
- [9] Khalid N, Pasha M, Ahmad H ,et al. Ontology Services Between Agents and OWL Based Web Services[J]. Semantics, Knowledge and Grid, Third International Conference, 2007(29-31): 176-181.
- [10] Kim P, Su-Kyoung. Implementation of Web Ontology for Semantic Web Application[J]. Advanced Language Processing and Web Information Technology, 2007. ALPIT 2007. Sixth International Conference, 2007(22-24):159-164.
- [11] Kim Dong-Soon, Hwang Suk-Hyung , Kim Hong-Gee. Concept Analysis of OWL Ontology Based on the Context Family Model[J]. Convergence Information Technology, 2007. International Conference, 2007(21-23): 896-901.

- [12] Choi K. IT Ontology and Semantic Technology[EB/OL].[2008-05-01] .http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?isnumber=4367992&arnumber=4368004&count=91&index=11.
- [13] Tho Q, Hui S, Forg A, et al. Automatic Fuzzy Ontology Generation for Semantic Web[J]. Knowledge and Data Engineering, 2006(6):842-856.
- [14] Chun Soon Ae, Geller J. Evaluating Ontologies Based on the Naturalness of Their Preferred Terms [C].Hawaii International Conference on System Sciences, 2008(7-10):238-238.
- [15] Samir T, Arpinar I B. Ontology Evaluation and Ranking Using OntoQA [J]. Semantic Computing, 2007. ICSC 2007. International Conference, 2007(17-19): 185-192.
- [16] Muhammad A, Wajahat N. Warnings for Disjoint Knowledge Omission in Ontologies[J]. Internet and Web Applications and Services, 2007. ICIW apos; 2007. Second International Conference, 2007(13-19): 43-45.
- [17] Martin H. Possible Ontologies: How Reality Constrains the Development of Relevant Ontologies [J]. Internet Computing, IEEE, 2007(1):90-96.

[作者简介]

杨良斌 男, 国际关系学院信息科技系讲师, 中国科学院国家科学图书馆博士生, 发表文章 3 篇。

黄国彬 男, 中国科学院国家科学图书馆博士生, 发表文章 40 余篇。

周静怡 女, 工作于中国科学院国家科学图书馆, 馆员, 发表文章 10 余篇。

[收稿日期: 2008-06-27]