

本体映射综述

王 顺 康达周 江东宇

(南京航空航天大学计算机科学与技术学院 南京 210016)

摘 要 作为不同本体之间知识共享和互操作的一种方法,本体映射受到越来越多的重视。根据本体映射过程将本体映射系统划分为五大功能组件,总结了本体映射系统中常用的相似度算法。通过梳理本体映射领域的最新发展成果,从不同层次、不同维度构建本体映射系统分类体系。介绍并比较了一些经典的本体映射系统,并对这些本体映射系统进行评价。最后指出本体映射将面临的挑战。

关键词 本体映射,本体,本体匹配

中图分类号 TP311 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.09.001

Survey of Ontology Mapping

WANG Shun KANG Da-zhou JIANG Dong-yu

(College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract As a method of knowledge sharing and the interoperability among different ontologies, ontology mapping has attracted more and more attention. According to the ontology mapping process, we divided ontology mapping system into five parts and summarized the similarity algorithm used in different ontology mapping systems. We groomed the status of art in the ontology mapping field of the recent years in this article, did deep analysis and comparison for a serial of the typical ontology mapping systems at different levels and dimensions. Introduction and comparison have been made on some classic ontology mapping systems, and these ontology mapping systems have been evaluated. The challenges of ontology mapping systems can be seen.

Keywords Ontology mapping, Ontology, Ontology matching

1 引言

2001年欧洲委员会启动了一个SWAP(Semantic Web and Peer-to-peer)的项目^[1],该项目需要在不同的本体之间进行互操作,而这些本体之间存在本体异构,这就为本体之间的互操作造成了障碍,从而促进了人们对本体映射问题的重视。建立本体的初衷就是为了知识共享,但是建立一个全局通用的大本体尚存在很大的困难,而现存的本体往往是由不同的本体开发团队构建的,本体构建标准不一致,从而造成了本体异构。由于本体异构的出现,要想完成信息交流的任务,就必须在本体之间架起语义映射的桥梁。最初这些映射过程都是手工完成的,但是随着语义Web的发展,在Web上用本体表示的信息越来越多,仅仅由人来完成这些工作已经力不从心,因而迫切需要找到一些方法来自动或者半自动地完成这种映射过程。

要想实现本体之间的映射,解决本体之间的知识共享问题,就目前建立全局的大本体面临的困难而言,本体映射^[2]就

是突破本体异构障碍的有效手段,是分布式环境下实现不同本体之间知识共享和互操作的基础性任务。

本文通过对已有的本体映射系统的分类和本体映射过程的分析,根据功能将本体映射划分为五大组件,并总结出本体映射过程中所用到的核心相似度算法。在此基础上,通过不同层次、不同维度对本体映射系统进行分类,总结出不同的分类下本体映射系统的特点,并对一些重要的本体映射系统进行详细介绍。最后对这些本体映射系统进行对比分析,并对不同分类下的本体映射系统的特点进行总结。

2 本体映射过程的分析

本体映射的分类在不同的文献中有不同的方法。Pavel Shvaiko^[3]将本体映射系统分为元素级和结构级的本体映射方法。其中元素级的本体映射方法包含着基于字符串、基于语言、基于约束的不同分类,而结构级的本体映射方法包含着基于图、基于类别的不同分类。Hooi^[4]将本体的映射方法分为术语映射、结构映射、语义映射、聚合映射。Noy^[5]将本体

到稿日期:2016-08-01 返修日期:2016-11-20 本文受江苏省基础研究计划(自然科学基金)资助项目(BK20130812),中央高校基本科研业务费专项资金(NS2014099),国防基础科研计划重点项目(JCKY2016206B001),武汉大学软件工程国家重点实验室开放基金(SKLSE2012-09-09)资助。

王 顺(1990-),男,硕士生,CCF会员,主要研究方向为语义网、本体论,E-mail:WANGSHUN611@163.com;康达周(1980-),男,博士,讲师,CCF会员,主要研究方向为语义网、本体论、描述逻辑;江东宇(1990-),男,硕士生,主要研究方向为语义网、本体论。

映射方法分为基于结构、基于机器学习和其他的本体映射方法。袁洋^[6]将本体映射系统分为模式级与实例级、元素粒度与实例粒度、基于语言与基于约束、匹配技术、辅助信息等。黄烟波等人^[7]将本体映射方法分为基于语法、基于实例、基于定义、基于层次结构的分类。这些本体映射系统的分类方法虽然不尽相同,但是本体映射系统的映射过程却基本相同。

2.1 本体映射功能组件和映射步骤的介绍

根据对本体映射过程的分析,可以将本体映射系统分为五大功能组件,分别为预处理组件、映射发现组件、资源组件、匹配器组件、映射表示组件。

预处理组件主要是消除参与映射的本体在表示上的差异(这些差异一般是语法、结构、语言等方面造成的异构),把待映射本体转换为同一语言的表示形式。

映射发现组件用于向匹配器组件输入将要进行匹配操作的映射候选对。

资源组件一般是为匹配器组件在进行匹配操作时提供相应的背景知识、领域约束等,在一些本体映射系统中该组件并不是必须具备的。

匹配器组件是本体映射最为核心的功能组件,不同的本体映射系统由不同匹配器组件构成。匹配器是计算概念相似

度的基本单元,综合先前的研究成果可以将这些匹配器分为词法匹配器、属性匹配器、实例匹配器、结构匹配器、权值分析器等。匹配器组件很少单独使用其中的某一种匹配器,更通用的做法是将不同的匹配器组合起来使用,这样可以显著提高映射的精度。

映射表示组件一般是将匹配器组件输出的映射结果以结构化的文本形式表示,作为本体映射系统最终的输出结果。

不同的映射系统并不需要完全具备这些功能组件,比如同一组织构建的领域本体映射系统有的就不需要预处理组件,因为进行映射的本体之间在概念的表示上并不存在太大的差异。但是通用的本体映射系统往往需要具备这些组件来完整地对本体之间的映射。

这些本体映射组件在具体的本体映射过程中的作用可以参考 Ehrig 与 Staab 对前人工作进行深入分析与总结后整理出的本体映射的 6 个基本步骤^[8]。Hooi^[4]将本体映射的过程进行了简单的概括,不同的本体映射系统虽然有着不同的功能设计,但整个映射过程大都符合这样的映射流程。图 1 所示的本体映射过程是根据 Hooi 对本体映射系统映射过程的描述并联系上述对本体映射系统每个映射功能的划分进行绘制的。

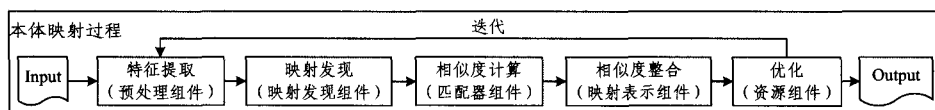


图 1 本体映射过程

下面对图 1 每个步骤中本体映射相应组件的作用进行说明。

步骤 1 特征提取。在本体中对概念名称、属性名称、概念实例等用于映射操作的各种特征的提取,一般在预处理组件中完成。

步骤 2 选取用于映射的概念对。在源本体和目标本体中分别选择概念进行映射的准备,一般在映射发现组件中实现。

步骤 3 计算选取的概念对特征的相似度。所选取概念之间的相似度计算是本体映射系统的核心步骤,一般是由匹配器组件中的各种匹配器来完成。

步骤 4 整合相似度。衡量实体间的相似度往往会使用多种不同的策略方法,从而产生多种相似度结果,要综合考虑每个相似度,从而得到一个合理的综合相似度,该操作一般在映射表示组件中完成。

步骤 5 优化。完成上一步操作后,已经获得待映射的每个实体之间的最初相似度,此时可以利用资源组件中的领域相关知识或者是运用启发式知识对结果进行调节,例如利用概念的结构之间的约束关系,若两个概念的子概念相似,则它们也可能相似。

步骤 6 迭代。反复进行步骤 1—步骤 5,直到得到满意结果。

2.2 本体映射中的粒度计算

本体规模的不断膨胀与本体描述语言的专业化给本体的复用造成了障碍。在本体映射中需要用到本体可能只是大本体中的一个本体片段,如果以本体实例和关系这些本体

基本元素作为映射单元,显得粒度太细,会产生语义缺失等问题;如果以本体库或者本体文档作为映射单元,又显得粒度太粗,缺乏本体映射的灵活性。在这些用于本体映射的基本映射元素中不存在显式的粒度描述方法。通过分析本体结构的特点,可以从中析取出有助于确定本体粒度的一些信息,有利于本体映射过程对映射粒度的计算。

在粒度计算理论中,粒度是指对信息颗粒大小的衡量,可以看作是从不同层次对问题进行的抽象与凝练。本体可以理解为特定领域相关概念的集合,以及其上的各种关系。通过借助于本体中概念的层次,定义的相似概念或者相关概念把概念分成若干集合,而这与粒计算模型的观点相似,说明本体可以描述为信息粒,这就为本体映射的粒度计算提供了理论依据和技术支持。

本体映射中粒度的计算依赖于粒计算理论。现有的粒计算模型可以大致分为基于模糊集理论和邻域观点的粒计算模型^[9]、基于粗糙集^[10]的粒计算模型^[11]、基于商空间的粒计算模型^[12]。从这些粒计算模型中可以看出,粒计算中最重要的是如何构造各种不同的粒,要构造粒就要按照一定的标准对论域进行划分,而不同的划分标准又对应不同的粒。本体作为描述概念与概念之间的概念模型,通过概念之间的关系来描述概念的语义,因此可以借鉴本体中的概念模型来确定不同的粒度。

在某些情况下可以采用概念层次树来表示粒度,因此将本体中的分类体系构造的概念层次结构应用于粒度的表示是可行的。本体的层次结构容易实现类的分解与合并,从而可以解决一些层次关系明显的问题。而层次关系不明显的问题

在构造这种层次结构时需要做出具体的定义。在本体中建立类之间的层次关系主要有以下两种方法。

(1)自顶向下法:由某一领域中最大的概念开始,然后逐步添加子类将这些概念细化,相当于粒度的分解。

(2)自底向上法:由最底层、最细小的类的定义开始,然后将这些细化的类组织在更加综合的概念之下,相当于粒度的合成。

将本体粒化得到的不同粒层上的同一个问题以不同的粒度、不同的细节表示,本体粒层之间的映射就建立了同一问题、不同细节描述之间的关系;进而在本体映射中利用本体粒之间的关系进行相应的相似度计算。

3 本体映射相似度计算方法的分类

本体映射最为核心的功能部件为匹配器,本体映射中用到的匹配器执行着本体映射最为核心的本体映射算法。在具体的映射过程中,不同的映射系统根据匹配器中所采用的相似度计算可以分为:编辑距离法(Edit Distance)、基于 WordNet 词典的相似度算法、松弛标记法、语义子图法、相似度传播法(Similarity Flooding)、向量距离法和 Jaccard 系数法等,这些方法从不同角度对进行本体映射的本体中含有的信息进行处理,从而得到本体不同粒度之间的相似度,以便于本体映射对的选择。这些相似度计算一般是在进行匹配操作的匹配器中进行计算的,对这些本体映射中的相似度计算方法进行分类,有助于根据每种相似度的算法特性深入理解每种本体映射系统的特性。

编辑距离法(Edit Distance)^[9-10]为字符串转换所需的最小数目的单元编辑操作,包括字符的插入、删除、替换及相邻字符的调换。编辑距离的计算方法简单,但是由于忽略了同形异义的字符串^[11]之间的语义信息,误差较大,容易产生错误映射结果或者丢失相似度的计算。这种方法在早期的本体映射系统中有很多应用。

基于 WordNet 词典的相似度算法^[12,17]作为一种语义字典,不同于传统的字典形式。在基于概念的本体映射中,WordNet 主要是根据概念之间的语义关系进行检索,将检索结果作为本体映射某阶段的输入来进行相应的计算。WordNet 还可以提供概念之间的上下位关系,根据这些上下位关系就可以推断出这些概念之间间隔的节点数,进而可以推断出概念之间的语义距离。WordNet 还可以在映射的本体之间的部分概念语义缺失的情况下对缺失语义信息的概念进行粗略的语义补充,方便后续映射。但是 WordNet 还是存在很多的局限性,如 WordNet 几乎没有关于句法的标注,虽然关于词语的概念够用,但是概念之间的关系仍然不足以进行推理。所以,WordNet 只是作为一种辅助的本体映射词典存在。

松弛标记法(Relaxation Labeling)^[18]是一种采用符号来描述特征的识别方法。在这种方法中,处理对象一般称为目标,而描述目标的符号则称为标记。松弛标记法首先对目标给定一组不确切的标记,然后通过迭代运算更新标记,最后求得这组目标的较为确切的标记集,算法的整个过程与人对某一事物的猜测推理过程相类似。该方法由于以迭代方法进行,因此易于实现,但其计算量太大,只有采用并行处理的

方法才可能充分发挥作用。因此,这种方法在近些年本体映射系统中的运用较少。

语义子图^[19]在本体映射的过程中虽然考虑到了元素的标识和注释在一定程度上反映了部分语义信息,但是本体的构建缺乏精确的统一标准,因此并非每个本体都会提供理想的描述文本。为了准确理解本体中元素的语义信息,往往还需要考虑与该元素相关的其他元素,这称为元素的语义上下文。借助本体图来分析元素的语义上下文,我们可以认为本体图中与元素相关的点和边构成了元素在本体中的语义上下文,这种语义上下文是本体图中的一个子图,称为元素的语义子图。利用语义子图的方法来对本体映射进行分析,是一种新的本体分析的方法。

相似度传播法(Similarity Flooding)^[20-21]基于已经获得的映射信息推算出相邻元素的相似度,以获得更多的映射对。相似度传播的思想是基于具有相似概念或子概念本身很可能相似。这种相似度计算的方法充分利用了本体概念之间结构上的关系。目前,此方法或者其改进方法已经被很多映射系统所采用,因为这种方法可运用本体概念结构上的约束关系明显地减少不必要的相似度的计算,从而减少本体映射系统的时间消耗。

向量距离法^[22-23]一般是将 RDF 和 OWL 的谓词作为语义参考点,然后利用对象之间语义参考点的连接把对象表示成向量,这样待比较的本体就被表示为相似度矩阵。在相似度计算的过程中,该方法采用图匹配算法来反复地更新矩阵中的数据,直到所有的数据都收敛为止。最终,本体之间任意两个概念的相似度都可以通过更新后的相似度矩阵查询得到。

Jaccard 系数法^[24]在信息检索或者是搜索引擎中经常使用到,在很多本体映射系统中的概念之间的相似度计算中也会使用到这种算法。这种方法用于集合之间的相似性,但是无法衡量度量集合差异值的具体值,只是获得类似于是否相同的结果,因此 Jaccard 系数法只关心集合之间共同具有的特征是否一致。该方法还有一个严重的缺陷,即未考虑集合中数据项出现的频率,仅仅统计了不同数据项出现的次数,当映射本体间富含的数据差异很大时,计算出的相似度结果很容易失真。

这些相似度计算方法在不同的本体映射系统中得到了应用。文献[25]中的 QOM 是一种本体映射框架,它的核心由相似度计算与合并模块、建立映射模块和迭代模块组成,其中字符串的相似度计算用到了编辑距离算法。文献[26]中的 MSBN 是一种基于多策略的贝叶斯网络的本体映射方法,它使用概念名称的编辑距离、概念的描述信息和实例特征来计算概念之间的相似度,最后使用本体的结构信息来辅助映射的查找。文献[27]中的 Cupid 在映射过程中对概念的名称信息、数据类型信息、约束信息以及元素所在的子结构信息进行相似度比较,并运用相似度传播算法作为本体映射的基础,这种单纯的基于概念的本体映射系统仅对本体中的概念进行映射处理,对本体实例中其他信息的利用较少。文献[28]中的 RiMOM 是一个动态多策略本体映射系统,在计算概念相似度时运用了编辑距离、WordNet 词典、向量距离、相似性传播等相似度算法。但是 RiMOM 在测试效率和稳定性方面难以

保证,在挖掘实例语义信息方面也有待改进。这种混合式本体映射系统代表了本体映射系统的发展趋势。文献[29]中的ASMOV使用概念的语言信息、内部结构信息、外部信息和个体信息,计算概念之间的相似度,多种信息的充分利用使得ASMOV能够获得比较可靠的本体映射结果。在计算概念之间的相似度时使用了WordNet词典,并且运用文本匹配算法计算词表距离。但是在基于大规模实例文档的实例匹配过程中,ASMOV还需要进一步优化。文献[30]中的WNPntoSim是一种混合型本体映射系统,它使用WordNet词典和原子概念标志来计算本体概念元素的结构语义相似度。文献[23]中的VBOM在本体向量化的过程中有效地使用数据类型属性和对象类型属性,为待比较的本体建立一个向量空间模型,然后提出一种权值分配机制,对概念和属性进行向量化处理,最后使用向量距离计算概念之间的相似度。文献[24]提出了一种综合概念相似度计算方法,利用Jaccard系数法来计算概念的实例相似度,但是这种方法未考虑实例个数的差异,当映射的本体之间的实例个数存在较大的差异时,得出的映射结果就不会太准确。

4 本体映射方法的分类

根据不同的本体映射系统的特点,从不同层次、不同维度对本体映射系统进行分析,并构建如图2所示的本体映射系统的分类体系。具体到每种本体映射系统,通过在不同的角度对它进行分类,使我们更加清晰地了解每种本体映射系统的映射领域、采用的映射技术和映射系统的特性。

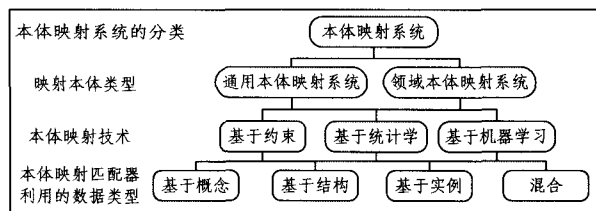


图2 本体映射系统的分类

通过对本体映射系统的应用范围进行研究,发现在不同时期开发的本体映射系统由于开发的目的不同,使得不同的本体映射系统会有不同的映射领域。可以将其分为两种^[31]: 1)通用本体的映射系统;2)领域本体的映射系统。

1)通用本体的映射系统^[32]通常采用不确定性推理、机器学习、相似度集合等技术,可满足多个应用领域的映射需求,在一般的映射环境下支持各种潜在的应用。通用本体映射系统为了对不同领域的本体进行映射,需要充分利用输入的本体中的各种信息。不同的领域本体映射系统可以利用各种领域知识进行约束,通用本体映射系统可用的信息一般是来自输入的本体以及本体所对应的实例。

2)领域本体的映射系统^[33]一般是针对领域问题的本体映射需求,使用具体的领域规则、启发式学习或者是背景知识进行辅助映射,在映射的过程中具有比较强的针对性。针对特定领域所开发的本体映射系统由于应用的领域已知,因此可以利用相应的背景知识在映射过程中进行一定的约束或者在创建映射规则时进行参考。不同于通用本体映射系统,领域本体映射系统在解决特定领域本体互操作方面具有优势。

根据本体映射的应用领域进行分类的方法主要是基于在

实际的应用中领域本体越来越多,而且领域本体之间的互操作需求越来越重要,因此不同领域本体之间的映射具有非常重要的现实意义。通常情况下,不同的领域之间进行低层次的互操作往往采用通用本体映射系统,因为通用的本体映射系统的应用范围广,没有领域约束的限制,操作灵活。但是为了应对具体应用领域中对本体之间的映射准确率的高要求,领域本体映射系统会根据具体的映射领域添加一些领域约束或者是利用一些知识库辅助本体映射,这种针对特定领域的辅助映射对提高领域本体映射的准确率有很大的帮助。

本体映射的技术发展至今,已经涌现出多种手段,而且随着新技术的不断发展,会有更多的新技术手段得到应用。通过对目前的本体映射技术进行分析,可以将本体映射系统划分为基于约束的本体映射方法、基于统计学的本体映射方法和基于机器学习的本体映射方法。

基于约束的本体映射方法^[34-35]是指在本体映射中定义了一些启发式规则,如“如果两个概念的子概念都相同,那么这两个概念是相似的”等。这些启发式规则是由领域专家手工定义的。其实这些规则的抽取来自于概念的定义和结构信息。在一些具体的应用领域,一些基于先验知识的领域约束信息就可以在本体映射过程中作为约束使用。

基于统计学的本体映射方法^[18,36]是指在映射过程中采用统计学中的方法。传统的基于实例的本体映射的方法简单地假定概念的各个数值属性是完全独立的,而没有考虑它们之间可能存在的对应概念依赖关系。事实上,实例中的每个数值属性取值分别与概念中的数值属性相对应,同时考虑到一个本体概念一般会描述多个本体实例,如果把概念中的每个数值属性都看成是离散变量,而一系列对应的实例数值属性取值则成为这些离散变量的值域空间,那么多个离散变量在值域空间的联合分布则成为识别该本体概念的一个重要特征,基于统计学的本体映射系统会充分利用这些统计特征进行本体之间的映射操作。

基于机器学习的方法^[37-38]是指在映射过程中采用机器学习的技术。机器学习的本体映射方法将映射问题转换成分类问题,为某个元素选择最优映射的问题就转换成对其进行分类的问题。通常利用已知映射结果的数据集训练分类器,当面临新的映射任务时,对每对元素是否匹配进行预测。基于机器学习的本体映射方法通常利用已有的机器学习方法,比如支持向量机、形式化概念分析、贝叶斯学习、神经网络等。

根据本体映射系统采用的映射技术进行分类是为了在具体的算法层面对本体映射过程进行分析,从而可以更加深入地理解每种本体映射系统的性能。基于约束的本体映射系统依赖于各种约束,这些约束往往是由专家来进行定义的,其好坏直接影响着本体映射系统的映射准确率。如果本体映射系统的应用领域发生变化,那么所有的约束条件可能会重新进行定义,这是一个非常复杂的过程,因此基于约束的本体映射系统在领域本体映射系统中应用得比较广泛。基于统计学的本体映射系统利用一些统计学中的方法,对进行本体映射的本体的一系列特征进行统计学上的分析,并构建相应的统计学模型进行分析。这种本体映射系统对源本体和目标本体的各类可以量化的统计指标有一定的要求,而且若参与本体映射的本体的规模越大,本体的映射越精确。基于机器学习的

本体映射系统则是利用最新的机器学习方法将本体之间的映射问题转换为分类的问题。机器学习方法的精确度依赖于训练集的质量和数量,训练集的数据量越大,数据的密度越大,训练出来的效果越好。大部分的机器学习方法被用于本体实例的映射过程中。

从本体映射的匹配器所执行的匹配操作的数据类型的角度,可以将本体映射方法分为基于概念的本体映射方法、基于结构的本体映射方法、基于实例的本体映射方法、混合式本体映射方法。

基于概念的本体映射方法^[39-41]一般是利用本体中的概念之间的相似度进行映射,在这种基于概念的本体映射的过程中往往需要对概念进行预处理,以消除概念之间在表达上的差异。本体之间基于概念的映射会将概念名称、概念描述、概念属性、概念之间的关系作为映射算法的输入。

基于结构的本体映射方法^[6]一般会利用本体概念之间在结构上的特殊约束关系或者是某个领域本体的特殊概念之间的关系进行约束,以对本体映射过程进行优化,或者是直接用于本体映射的核心映射的全过程。

基于实例的本体映射方法^[42-45],其输入是两个本体的所有元素及其对应的实例。在本体库中每个元素都有多个实例,每个实例通常包含实例名称以及一组相关的属性值。在以往的实验中,当实例的文本较长时,基于实例的本体映射方法的效果较好;当实例的文本较短或者是单词分布过于稀疏时,其效果比较差。

混合式本体映射方法^[46-47]主要是运用多种本体映射手段,然后对每种手段得出的最终映射结果进行综合处理,从而得到最终的映射结果。这种方法能够充分利用不同的匹配器在本体映射过程中的筛选,找到最佳的映射对,但是其时间和空间开销很大。这种方法主要是应用于本体中拥有大量信息的情况,一般是通过相关算法,采用名称和实例等文本信息建立一系列的学习模型,进而得到本体概念之间的初步相似度,然后再通过启发式方法^[35]或者是一系列的领域约束来进行不断的迭代,最终完成本体之间的映射过程。近年来,单纯地采用一种匹配器的本体映射系统比较少见,更多的是在映射过程中利用多种数据类型匹配器组合的混合式本体映射方法^[24,48-53]。

根据本体映射系统匹配器中操作的数据类型进行本体映射系统的分类主要是不同的本体映射系统由于操作的本体数据类型不同,需要针对这些数据类型进行匹配器的设计。从本体映射系统在不同的测试集上的表现来看,基于概念的本体映射方法大多应用于领域本体之间进行映射的情况,而且其常常需要外部知识库作为扩展的资源使用。在处理一些本体映射时,可能还需要对进行映射的本体进行预处理操作,以消除概念之间在表现形式上的差异。基于结构的本体映射方法一般应用于通用本体之间的映射,利用本体概念之间的结构关系进行相似度传递或者利用结构约束关系对本体概念之间的相似度进行计算,这种映射方法便于进行自动化操作,在增加一些人为的领域约束的情况下映射准确率较高,但是由于在处理结构信息时需要花费大量的时间和空间,目前比较好的通行做法是采用并行化技术来减小映射系统的时间和空间开销。基于实例的本体映射系统的映射准确率依赖于实例

信息的数量,实例的数据越多,系统的映射准确率就越高。基于实例的本体映射系统一般采用机器学习的方法对不同的实例集进行机器学习,实例集的划分需要手工完成,在划分实例集的过程中往往需要根据映射要求添加一系列的约束条件,约束条件的制定对机器学习的训练集的质量有着很大影响。混合式本体映射方法是当今比较流行的映射系统普遍采用的一种映射方法,这种方法在映射的核心部分运用多种映射方法来分别或者并行进行输入本体之间的映射,然后再将每种方法得到的映射结果进行综合处理,从而得出最后的映射结果。这种混合式本体映射系统由于综合了各种本体系统之间的优点,因此比单纯采用某种映射方法的系统的准确率高。但是这种混合式本体映射系统对不同的本体之间的映射没有专用的映射方案,如缺乏实例信息本体之间的映射就不需要基于实例的本体映射过程,改进的方法是添加动态多策略的本体映射算法来自动进行映射策略的选择。单纯地利用某一种类型数据的本体映射系统已经比较少见,比较流行的做法是对不同数据类型的匹配器进行匹配操作,最后得出映射结果。这种综合利用本体中概念、结构、实例数据的本体映射系统具有比较稳定的映射准确率。

5 本体映射系统的介绍和比较

5.1 经典本体映射系统的介绍

近些年出现了一些在通用的本体数据集上有较好表现的本体映射系统^[54-56],尤其是国际语义网会议举办的 OAEI (The Ontology Alignment Evaluation Initiative) 竞赛中所展示的本体映射系统反映了当前本体映射最新的发展成果,具有较高的映射准确率。本文对在 2015 年 OAEI 竞赛中表现比较突出的两种本体映射系统 AML 和 Lily 进行介绍。

5.1.1 AML

Daniel Faria 在 2015 年的 OAEI 中参赛的 AML 本体映射系统就是基于本体概念的本体映射系统^[57-58]。AML 本体映射系统的映射过程如图 3 所示。匹配器中包含了基本匹配器、背景知识匹配器、单词或字符串匹配器、结构匹配器、属性匹配器,其中的结构匹配器只对小、中型本体起作用,利用本体中关系的继承进行相似性的传递。

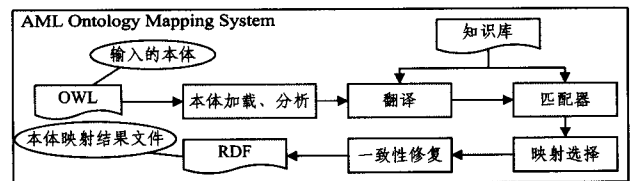


图3 AML本体映射系统

在 AML 本体映射系统中,本体加载和分析主要是应用 OWL API 读取输入的本体,并检索必要的信息来填充到相应的数据结构中。在对用不同语言创建的本体进行翻译和对输入本体之间进行匹配操作时,AML 运用了背景知识作为知识库辅助本体的翻译和匹配过程。AML 注重通过复用和使用一些构件来进行知识积累,其中包含一些新颖的构件,如一致性修复模型和自动化背景知识资源选择算法等。由于用于竞赛的本体规模越来越大,并且在这种环境下还要处理用户的反馈信息,这对当前的 AML 本体映射系统来说是一个很大的挑战。

在 OAEI 的标准映射本体 bibliography ontology 中, AML 系统获得了 99% 的准确率, 近乎完美, 但其 F 度量值仅 57%, 召回率仅 40%。AML 本体匹配系统并没有针对实例数据做一些相应的处理, 而这些实例数据却是一些基于实例进行本体映射系统的映射依据, 因此 AML 还有很大的改进空间。另外, 由于在进行本体映射的过程中使用了多种匹配器进行匹配操作, 但是不同的本体之间适用的匹配器不尽相同, 而 AML 缺乏对匹配器的动态设置, 因此会导致一些不必要的匹配器也进行匹配操作, 从而浪费一定的时间。相对于 RiMOM^[28] 这类动态多策略本体映射系统来说, AML 在匹配器的动态设置上便显得不太灵活。

5.1.2 Lily

中国东南大学的 Lily 本体映射系统^[59-61] 在近几年 OAEI 竞赛中的表现比较抢眼。Lily 在大规模本体映射中为了减小时间和空间复杂度, 使用了特殊的应对策略。在基于结构的本体映射中, 相似度的正确传递可以有效减小映射过程的时间和空间开销, 但是相似度的错误传递会导致更多错误的出现。为了避免这些情况, Lily 运用了特殊的处理方法来规避不必要的相似度传递。Lily 系统通过本体实体之间的结构关系进行一系列的筛选过程, 避免了源本体和目标本体之间不必要的映射操作。其基本原理总结如下:

(1) 如果两个概念(A 和 B)之间有很高的相似度, 那么概念 A 的子概念和概念 B 的父概念之间的相似度计算就可以跳过(反之亦然)。

(2) 如果两个概念(A 和 B)之间的相似度很低, 那么利用本体映射过程中的相邻概念之间的语义局部现象就可以跳过概念 A 的邻接概念与 B 概念之间的相似度计算。

在 2015 年的 OAEI 竞赛本体中的标准测试集中, Lily 系统获得了 97% 的准确率, 召回率达到了 83%, F 度量值为 90%。但是在大规模本体映射的时间和空间开销方面, Lily 的表现一般, 其虽然在大型本体的映射过程中通过一些处理方法避免了不必要的时空开销, 但是在真正的映射操作上与传统的映射手段相比却没有太多的改进。与 AML 本体映射系统相比, Lily 对本体之间在结构上的关系的充分利用对其映射精度有很大的帮助, 但是在映射过程中对时间和空间的开销还需要进行优化。

5.1.3 GLUE

GLUE^[49] 系统是由 Anhai Doan 等人提出的一种典型的混合式本体映射系统, 其体系结构如图 4 所示。

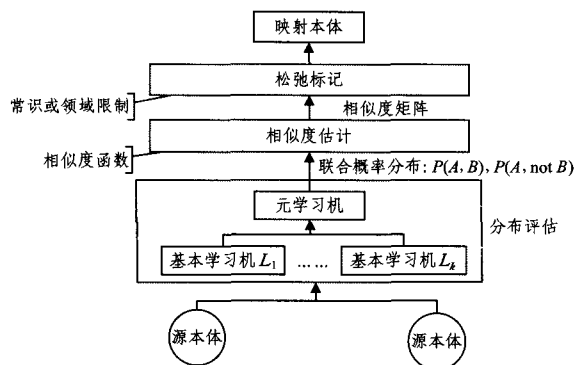


图 4 GLUE 体系结构

GLUE 的映射机制可以概括如下:

(1) 不同类型的信息能够在本体的匹配过程中有不同的贡献。GLUE 使用不同的学习机的集合, 并利用这些学习机做出的预测构建元学习机。

(2) 领域约束的利用。GLUE 使用一些领域约束和启发式方法来提高本体匹配的精度。

(3) 将从状态分布评估器中得到一系列概率向量直接输入后面的相似度评估器。结合用户给出的相似度评估函数, 相似度评估器对概念之间的相似度进行计算, 进而得到两个本体之间的相似度矩阵。

(4) 相似度矩阵被直接输入到标志修改器中, 并结合领域约束和启发式规则进行迭代, 直到找到最符合领域约束和公共知识的最优匹配和映射。

GLUE 在不同的测试集上的映射精度为 66%~97%, 这说明 GLUE 映射系统没有很好的映射稳定性。在缺乏足够实例的情况下, 基于实例的机器学习很难得到有效的映射信息。在本体的映射过程中, 本体实体中的关系和属性并没有得到利用, 这是 GLUE 需要改进的地方。但是不可否认, GLUE 系统为本体映射的方法起到了一定的推动作用, 尤其是各种学习器的组合映射在之后出现的一些本体映射系统中得到了进一步的优化, 这种混合式本体映射系统的架构也成为当今大多数本体映射系统的基本参照。但是在本体映射中的匹配器的功能更加强大, 映射的精度也更加稳定且自动化程度更高, 同时可利用的背景知识也更加丰富。

5.2 主流本体映射系统的介绍

近些年来基于多种技术手段的本体映射系统层出不穷, 这些本体映射系统由于在一些技术领域上有所创新而受到人们的关注。在大量本体映射需求的背景下, 有些本体映射系统在不断的技术更新中逐渐成为主流的本体映射系统。基于本文对本体映射方法的分类, 下面选取了比较具有代表性的采用主流技术的本体映射系统进行着重介绍。

5.2.1 HowNet 与 CCD 本体映射方法

北京大学的向春丞曾在 2013 年初探索了知网(HowNet)与中文概念辞书(Chinese Concept Dictionary, CCD)两部词典的映射方法^[62]。首先利用两者的词汇信息、语言信息以及语义信息定义映射特征, 然后给出将样本实例集划分为正例集、负例集以及测试集的策略, 接着利用机器学习技术发现映射关系, 最后考虑本体的整体结构信息并利用相似度传播算法对初始映射结果进行调整。实验表明, 最终概念之间的一对一和一对多的映射关系的准确率可达到 94% 和 87.5%。

作为基于实例的本体映射系统, HowNet 和 CCD 中名词概念之间的映射中, 参与映射的本体中实例的概念粒度以及属性之间存在较大差异, 导致部分映射未能有效进行。其次, 该映射系统对实例样本集的限制过于复杂, 很难有太大的扩展性, 测试的实例样本集的规模也受到了限制。最后, 其并未充分考虑时间和空间的开销, 本体映射的算法也需要进行优化。与 Lily 和 AML 相比, 该映射系统仅依赖于本体的实例信息进行本体映射, 实例信息的多少对映射结果有很大的影响, 而且该本体映射系统的应用领域很难有大的拓展。

5.2.2 OM-Matchers

张凌宇等人^[63]于2015年提出了一种基于多种匹配器的本体映射方法(OM-Matchers)。OM-Matchers采用名称匹配器、内容匹配器、实例匹配器和结果匹配器,分别计算概念之间的不同类型信息之间的相似度。其次,OM-Matchers根据不同类型信息占总信息量的比重为这些匹配器分配权重,并为概念对计算一个最终的相似度值。最后,OM-Matchers采用迭代的映射算法为相似度大于给定阈值的概念对建立映射关系。在OAEI竞赛的标准测试集的部分数据上进行验证实验,实验结果的准确率达到了90%,召回率为91%,F度量值为90%。实验表明,OM-Matchers可以有效利用本体所包含的多种类型的信息,并精确计算概念之间的相似度;而且其所采用的迭代映射策略可以反复利用多种类型的匹配器来计算概念之间的相似度,从而提高了映射的准确率、召回率和F度量值。但是,作为一种通用的本体映射系统,OM-Matchers未能有效实现更大规模的本体映射任务,通过反复利用多种匹配器来提高本体映射的准确度会有局限性,而且在本体映射系统的设计过程中未能充分考虑该系统的时间和空间利用率。

5.2.3 RiMOM

RiMOM(Risk Minimization based Ontology Mapping)^[64]是多策略动态本体映射系统,基于贝叶斯决策理论,将本体映射作为最优决策的发现问题的,从而提出基于风险最小化的本体映射模型。决策行为定义为所有可能的映射,因此映射最

优化问题就转换为决策行为最优化问题,即搜索风险最小的决策行为。采用基于约束的本体映射技术,在具体的映射策略上采用实例、概念、结构的混合式本体映射策略。该方法不仅在1:1的映射上取得了较好的效果,而且还实现了n:1映射。

实验在两个测试集 Course Catalog I 和 Company Profile 上进行。Course Catalog I 数据集中的本体分别描述康奈尔大学和华盛顿大学的课程信息,两个本体的概念和属性的名称定义比较相似。Company Profile 数据集中的本体分别描述了Yahho.com和The Standard.com公司的商务信息。这两个数据集都分别包含有两个异构本体,因此映射任务是发现两个本体相互之间的映射关系。对RiMOM在这两个数据集上的映射结果进行分析可以发现,查准率为73%~91.6%,查全率为83.44%~92%,表明了RiMOM本体映射系统的有效性。通过与传统的GLUE系统进行对比实验,发现RiMOM本体映射系统在映射查准率上提高了10%,说明RiMOM的映射方法明显优于GLUE中采用的映射方法。RiMOM的缺陷是没有充分利用领域知识和启发式知识,系统中的合并策略和启发式规则还不足以有效支持映射过程。

5.3 本体映射系统之间的比较

根据前面对本体映射系统的介绍,本小节对这些本体映射系统进行比较,具体内容如表1所列。根据本文对本体映射系统的不同应用领域、不同本体映射方法、不同本体映射系统匹配器处理数据类型的3种分类,对这些本体映射系统进行区分。

表1 不同映射系统的比较

本体映射系统的特征	AML	Lily	HowNet 与 CCD 映射系统	GLUE	OM-Matchers	RiMOM
映射的本体类型	领域本体	通用本体	领域本体	通用本体	通用本体	通用本体
本体映射技术	基于约束	基于约束	基于约束、机器学习	基于统计学、机器学习	基于约束、统计学	基于约束
匹配操作的数据类型	混合	基于结构	基于实例	基于概念	混合	混合
映射本体基数	n:m	1:1	1:1/1:n	1:1/1:n	1:1	1:1/1:n
核心映射算法	背景知识 选择算法	相似度 传播算法	机器学习、相似度传播算法	机器学习、相似度矩阵算法	Jaccard 系数法	相似度传播算法
预处理程度	较好	一般	一般	一般	一般	较好
自动化程度	自动	半自动	半自动	半自动	自动	半自动
背景知识	需要	—	需要	需要	—	—
用户反馈	支持	—	—	—	—	—
辅助映射的数据	知识库	—	—	—	重用映射结果	—
可扩展性	强	一般	一般	一般	较强	一般

下面对在本体映射系统的比较中需要用到的但前文未提到的一些概念进行解释。

(1)匹配基数:本体映射会产生6种本体映射关系,分别是1:1,1:n,n:1,n:m,1:null和null:1。因为n:1是1:n的反向操作,所以本文未做区分。大多数的本体映射方法只会考虑1:1的本体映射关系,因为n:m的映射关系需要在匹配算法中采用更加复杂的操作。

(2)预处理的程度:现在大多数的本体映射系统会在本体映射之前对输入的源本体和目标本体进行预处理操作,这些操作包括分词、停用词过滤、词干提取等。根据预处理程度的不同,将包含分词、停用词过滤和词干提取的预处理定义为高程度的预处理,将不完全包含这3种操作的预处理定义为一般程度的预处理。

(3)辅助映射的数据:大多数的匹配方法不仅依赖于所输

入的本体信息,还会用到一些辅助信息,例如字典、以前的匹配结果及用户的反馈等。

(4)可扩展性:可扩展性表现在一些本体映射系统会根据本体映射的具体应用领域,在匹配组件中加载与具体的领域相关的匹配器,这样的本体映射系统才更加灵活。可扩展性较强的本体映射系统对具体的领域具有较强的适应性和稳定性。

(5)自动化程度:映射系统的自动化根据本体映射系统在进行映射的过程中手工参与的程度进行划分。忽略在本体映射系统中预处理组件阶段对源本体和目标本体输入到映射系统的手工操作,在其他的本体映射阶段没有手工干预的本体映射系统被认为是全自动的本体映射系统,在其他的本体映射阶段有手工干预的本体映射系统称为半自动的本体映射系统。

从表1中不同映射系统的比较和上文对不同分类下的本

体映射系统的特点可以看出, AML 本体映射系统由于对输入源本体和目标本体进行了较高级别的预处理过程, 因此在映射过程中可以实现更高级别的自动化; 而且在匹配器中对数据的利用上运用了混合式的方法, 这种方法使多种数据参与到本体映射的过程, 并且支持用户对映射结果进行反馈, 使得系统具有较高的映射准确率。但是其在匹配组件中缺乏动态的匹配器选择策略, 使得映射操作比其他几种本体映射系统耗时。与 AML 本体映射系统相比, Lily 对本体之间在结构上的关系的充分利用对其映射精度有很大的帮助, 在映射过程中采用基于约束和基于结构的本体映射方法, 使得相似度传播算法可以充分利用结构上的约束关系进行本体映射操作。但是 Lily 在开始阶段添加了锚点的操作, 在映射过程中对时间和空间的开销还需要进行优化。GLUE 采用了基于统计学和机器学习的本体映射方法, 在映射的过程中通过不同的学习机对实例数据进行利用, 但是缺乏对其他数据的有效处理, 使其可扩展性一般, 不能根据参与映射的本体含有实例的多少进行映射策略的选择, 也使其映射准确率不稳定。HowNet 与 CCD 映射系统匹配器操作的数据类型是基于实例的, 在映射技术上采用基于约束的方法, 但是该映射系统对实例样本的限制太过复杂。与 AML 和 Lily 相比, HowNet 与 CCD 映射系统未能充分利用本体中其他有效的信息, 仅依赖于实例信息, 可扩展性一般。OM-Matchers 本体映射系统采用在匹配操作中可以利用多种数据的混合式映射方法, 而且有效的预处理的过程很好地支持了系统自动化的实现, 在辅助映射数据上支持重用映射结果, 使其可扩展性较强。Ri-MOM 基于相似度传播理论提供了 3 种相似度聚合策略, 分别是概念-概念、概念-属性、属性-属性聚合方法。在此基础上, 系统通过多个启发式规则剔除不可靠的映射, 精炼匹配结果。

5.4 未来研究方向的探讨

由于现有的本体大多采用专门的本体描述语言, 且容量巨大、内容复杂, 基于约束的和统计学的本体映射系统由于缺乏灵活的功能扩展, 在设计具体的本体映射系统时难以有效应对未来映射数据的变化。随着本体规模的扩大, 基于约束和统计学的本体映射技术的本体映射系统的时间复杂度和空间复杂度会明显增大, 最终达到无法满足应用要求的程度。到目前为止, 仍然存在几个亟需解决的问题, 首先目前的方法能够处理的映射类型十分有限, 大多数研究都集中于 1:1 的映射, 然而统计表明了现实世界的异构数据大约有 22%~25% 的映射都不是这种简单的映射类型; 其次, 本体包含多种类型的信息, 但利用了所有可用信息的系统还是比较少。总体来讲, 本体映射还是缺少一个更好的理论模型, 以便于综合利用所有的本体信息进行映射发现, 同时支持多种类型的映射发现。

根据对现有的本体映射系统的分析, 从本体映射系统采用的映射技术的层面来说, 采用基于机器学习的本体映射技术的本体映射系统越来越多。这是由于现有计算机软、硬件的持续发展和机器学习算法的持续改进, 使得机器学习对大量数据的处理比传统的数据处理手段高效得多。机器学习算法的高效性和精确性基于训练集的质量和数量, 机器学习训

练集的数据量越大, 数据的密度越大, 训练的效果越好。因此, 基于机器学习的本体映射技术较好地解决了由于本体规模不断膨胀给本体映射所带来的瓶颈问题, 这也是未来本体映射技术的发展趋势。

结束语 本体异构的出现导致了不同本体之间的知识共享和互操作面临困难, 本体映射为解决本体异构问题提供了新思路。现有的本体映射系统由于应用的领域、采用的技术手段和映射过程中利用的数据类型不同, 因此缺乏有效的分类体系。本文根据已有的分类方法, 构建出基于不同层次、不同维度的本体映射体系, 为本体映射系统提供了分类方法。根据本文构建的本体映射系统的分类方法, 对近些年来比较经典的本体映射系统进行介绍和比较。

近些年来, 本体的规模随着应用领域数据产生的速度急剧膨胀, 大数据的产生导致大本体的产生, 但是现在的本体映射系统在大数据环境下缺乏有效的应对手段, 这些问题是本体映射将要面临的挑战和需要解决的主要问题。

参考文献

- [1] SWAP [EB/OL]. [2010-10-15]. <http://swap.semanticweb.org/public/in-dex.htm>.
- [2] WONG A K Y, RAY P, PARAMESWARAN N, et al. Ontology mapping for the interoperability problem in network management[J]. IEEE Journal on Selected Areas in Communications, 2005, 23(10): 2058-2068.
- [3] SHVAIKO P, EUZENAT J. A Survey of Schema-Based Matching Approaches[J]. Lecture Notes in Computer Science, 2005, 4(1): 146-171.
- [4] HOUI Y K, HASSAN M F, SHARIFF A M. A Survey on Ontology Mapping Techniques [M] // Advances in Computer Science and its Applications. Springer Berlin Heidelberg, 2014: 15-33.
- [5] NOY N F. Ontology Mapping [M] // STAAB S, STUDER R. Handbook on ontologies. Springer Berlin Heidelberg, 2009: 573-590.
- [6] YUAN Y, LI S P. The Overview of Ontology Integration Approaches Based on Semantic Web [J]. Computer Science, 2004, 31(5): 5-8. (in Chinese)
袁洋, 李善平. 基于语义 Web 的本体映射方法综述 [J]. 计算机科学, 2004, 31(5): 5-8.
- [7] HUANG Y B, ZHANG H Y, LI J H, et al. Survey of Ontology Mapping Approaches [J]. Computer Engineering and Applications, 2005, 41(18): 27-29. (in Chinese)
黄烟波, 张红宇, 李建华, 等. 本体映射方法研究 [J]. 计算机工程与应用, 2005, 41(18): 27-29.
- [8] EHRIG M, STAAB S. QOM-Quick Ontology Mapping [J]. Lecture Notes in Computer Science, 2004, 3298: 356-361.
- [9] GRANULARITY I. Fuzzy Sets and Information Granularity [J]. Fuzzy Sets Fuzzy Logic & Fuzzy Systems, 1996, 6: 433-448.
- [10] PAWLAK Z. Rough sets [J]. International Journal of Parallel Programming, 1982, 11(5): 341-356.
- [11] LIU Q, LIU Q. Granules and Applications of Granular Compu-

- ting in Logical Reasoning[J]. Journal of Computer Research and Development, 2004, 41(4): 546-551. (in Chinese)
- 刘清, 刘群. 粒及粒计算在逻辑推理中的应用[J]. 计算机研究与发展, 2004, 41(4): 546-551.
- [12] ZHANG L, ZHANG B. The quotient space model of qualitative reasoning(II)[J]. Journal of Anqing Teachers College (Natural Science Edition), 1990(1): 15-20. (in Chinese)
- 张铃, 张钺. 定性推理的商空间模型(II)[J]. 安庆师范学院学报(自然科学版), 1990(1): 15-20.
- [13] BAX E T, SWETT I D. Edit distance string search; US, US758 4173[P]. 2009.
- [14] CORMODE G, MUTHUKRISHNAN S. The string edit distance matching problem with moves[J]. ACM Transactions on Algorithms, 2007, 3(1): 1-11.
- [15] CHEATHAM M, HITZLER P. String similarity metrics for ontology alignment[M]. The Semantic Web-ISWC 2013. Springer Berlin Heidelberg, 2013: 294-309.
- [16] REYANUD C, SAFAR B, REYNAUD C, et al. Exploiting WordNet as Background Knowledge[C]// International Workshop on Ontology Matching. 2007.
- [17] WU Z, PALMER M. Verb Semantics and Lexical Selection [C]// ACL Proceedings of Annual Meeting on Association for Computational Linguistics. 1995: 133-138.
- [18] DOAN A H, MADHAVAN J, DOMINGOS P, et al. Learning to map between ontologies on the semantic Web [C]// WWW 2002. 2002: 662-673.
- [19] WANG P. Ontology Matching Method Based On Semantic Subgraph[C]// 2008 China National Computer Congress. 2008. (in Chinese)
- 汪鹏. 基于语义子图的本体匹配方法[C]// 2008 中国计算机大会. 2008.
- [20] MLNIK S, GARCIA-MOLINA H, RAHMA E. Similarity Flooding: A Versatile Graph Matching Algorithm and Its Application to Schema Matching[C]// Proceedings of the 18th International Conference on Data Engineering. IEEE Computer Society, 2015: 117-128.
- [21] MARSHALL B, CHEN H, MADUSUDAN T. Matching knowledge elements in concept maps using a similarity flooding algorithm[J]. Decision Support Systems, 2006, 42(3): 1290-1306.
- [22] TOUS R, DELGADO J. A Vector Space Model for Semantic Similarity Calculation and OWL Ontology Alignment[J]. Lecture Notes in Computer Science, 2006, 4080: 307-316.
- [23] EIDOOON Z, YAZDANI N, OROUMCHIAN F. Ontology Matching Using Vector Space[M]// Advances in Information Retrieval. Springer Berlin Heidelberg, 2008: 472-481.
- [24] CAO Z W, QIAN J, ZHANG W M, et al. A Compositive Approach for Concept Similarity Computation[J]. Computer Science, 2007, 34(3): 174-175. (in Chinese)
- 曹泽文, 钱杰, 张维明, 等. 一种综合的概念相似度计算方法[J]. 计算机科学, 2007, 34(3): 174-175.
- [25] EHRIG M, STAAB S. QOM-Quick Ontology Mapping[M]// The Semantic Web-ISWC 2004. Springer Berlin Heidelberg, 2004: 683-697.
- [26] ZHANG L Y, MA Z M, YAN L. Ontology Mapping Based on Multi-strategy and Bayesian Network [J]. Journal of Chinese Computer Systems, 2011, 33(11): 2385-2391. (in Chinese)
- 张凌云, 马宗民, 严丽. 一种基于贝叶斯网络模型及多策略计算的本体映射方法[J]. 小型微型计算机系统, 2011, 33(11): 2385-2391.
- [27] MADHAVAN J, BERNSTEIN P, RAHM E. Generic schema matching with cupid[C]// Proceedings of the International Conference on Very Large Databases (VLDB). Berlin: Springer, 2001: 49-58.
- [28] LI J, TANG J, LI Y, et al. RiMOM: A Dynamic Multistrategy Ontology Alignment Framework [J]. IEEE Transactions on Knowledge & Data Engineering, 2008, 21(8): 1218-1232.
- [29] SHIRONOSHITA E P, KABUKA M R. ASMOV: results for OAEI 2010 [C]// International Conference on Ontology Matching. 2010: 126-133.
- [30] HE W, YANG X, HUANG D. A Hybrid Approach for Measuring Semantic Similarity between Ontologies Based on WordNet [C]// Knowledge Science, Engineering and Management, International Conference (KSEM 2011). Irvine, CA, USA, 2011: 73-77.
- [31] TAO J, SUN T. A Survey on Model for Ontology Mapping Systems Since 2009[J]. Library and Information Service, 2011, 55(12): 101-104. (in Chinese)
- 陶俊, 孙坦. 2009 年以来本体映射系统模型研究综述[J]. 图书情报工作, 2011, 55(12): 101-104.
- [32] YE F B, TANG R Z, WANG Z X, et al. Ontology mapping transform technology based on general ontology[J]. Journal of Zhejiang University (Science), 2009, 43(12): 2274-2280. (in Chinese)
- 叶范波, 唐任仲, 王正肖, 等. 基于通用本体的本体映射变换技术[J]. 浙江大学学报(工学版), 2009, 43(12): 2274-2280.
- [33] CHEN H, CHEN N J, SONG Y. A Reserarch on Ontology Mapping Frame[J]. Journal of Hunan University of Science and Engineering, 2009, 30(12): 58-62. (in Chinese)
- 陈宏, 胡宁静, 宋艳. 领域本体映射框架探讨[J]. 湖南科技学院学报, 2009, 30(12): 58-62.
- [34] LI H, SU L. Comprehensive method of computing ontology similarity based on association rules[J]. Journal of Computer Applications, 2012, 32(9): 2472-2475. (in Chinese)
- 李华, 苏乐. 基于关联规则的本体相似度综合计算方法[J]. 计算机应用, 2012, 32(9): 2472-2475.
- [35] LI J F, HU W. Discovering Complex Semantic Mappings as Learning Rules Across Ontologies [J]. Sciencepaper Online, 2011, 4(20): 1833-1838. (in Chinese)
- 陈建锋, 胡伟. 一种基于规则学习的本体映射方法[J]. 中国科技论文在线, 2011, 4(20): 1833-1838.
- [36] QIAN P F. Mapping & Merging of Heterogeneous Ontologies Based on Fuzzy UZZY Similarity and Verification[D]. Shanghai: Shanghai Jiao Tong University, 2008. (in Chinese)
- 钱鹏飞. 基于模糊相似度的异构本体映射、合并及校验方法的研究[D]. 上海: 上海交通大学, 2008.
- [37] MIAO Y C. Application of Artificial Intelligence in Ontology

- Matching[J]. *Ship Electronic Engineering*, 2015(7): 71-74. (in Chinese)
- 苗永昌. 人工智能在本地映射中的应用[J]. *舰船电子工程*, 2015(7): 71-74.
- [38] SPOHR D, HOLLINK L, CIMIANO P. A Machine Learning Approach to Multilingual and Cross-Lingual Ontology Matching [J]. *Semantic Web-Iswe 2011, Pt 1*, 2011, 7031: 665-680.
- [39] WANG F, CHEN J. Research on Multi-strategy Ontology Mapping Based on Conceptual Similarity Computing[J]. *Computer Technology and Development*, 2015(4): 38-42. (in Chinese)
- 王凡, 陈健. 基于概念相似度计算的多策略本体映射研究[J]. *计算机技术与发展*, 2015(4): 38-42.
- [40] SANTOS E, FARIA D, PESQUITA C. Ontology alignment repair through modularization and confidence-based heuristics[J]. *Plos One*, 2013, 10(12): e0144807.
- [41] CHENG J H, ZHENG S H, LI W L, et al. Weight Determine Method for Comprehensive Similarity Calculation about Concept of Ontology[J]. *Journal of Jilin University (Science Edition)*, 2014(6): 1272-1276. (in Chinese)
- 成锦晖, 郑山红, 李万龙, 等. 本体领域综合概念相似度计算中的权重确定方法[J]. *吉林大学学报: 理学版*, 2014(6): 1272-1276.
- [42] ZHONG Z C, ZHONG Y, GUO G. Research on ontology mapping method based on instances[J]. *Information Technology*, 2010, 34(8): 22-26. (in Chinese)
- 仲志成, 仲颖, 郭刚. 基于实例的本体映射方法研究[J]. *信息技术*, 2010, 34(8): 22-26.
- [43] CASTANO S, FERRARA A, MONTANELLI S, et al. Ontology and instance matching[M]// *Knowledge-Driven Multimedia Information Extraction and Ontology Evolution*. Springer Berlin Heidelberg, 2011: 167-195.
- [44] KHIAT A, BENAÏSSA M. Approach for Instance-Based Ontology Alignment: Using Argument and Event Structures of Generative Lexicon[J]. *Metadata and Semantics Research*, 2014, 478(2): 120-127.
- [45] SCHOPMAN B, WANG S, ISAAC A, et al. Instance-Based Ontology Matching by Instance Enrichment[J]. *Journal on Data Semantics*, 2012, 1(4): 219-236.
- [46] JIANG Z, YAO X M, LIN L F. Feature-based adaptive method of ontology mapping[J]. *Journal of Zhejiang University (SCIENCE)*, 2014, 48(1): 76-84. (in Chinese)
- 蒋湛, 姚晓明, 林兰芬. 基于特征自适应的本体映射方法[J]. *浙江大学学报(工学版)*, 2014, 48(1): 76-84.
- [47] FANG A, HONG N, WU S. An Integrated Biomedical Ontology Mapping Strategy Based on Multiple Mapping Methods[M]// *Web Information Systems Engineering-WISE 2013 Workshops*. Springer Berlin Heidelberg, 2013: 373-386.
- [48] DO H H, RAHM E. COMA: a system for flexible combination of schema matching approaches[C]// *Proceedings of the 28th International Conference on Very Large Data Bases*. 2002: 610-621.
- [49] CHOI N, SONG I Y, HAN H. A Survey on Ontology Mapping [J]. *SIGMOD Record*, 2006, 35(3): 34-41.
- [50] YU X, WU J. A New Ontology Optimization Algorithm for Similarity Measuring and Ontology Mapping in Multi-dividing Setting[J]. *Journal of Computational Information Systems*, 2015, 11(9): 3297-3305.
- [51] ANAM S, YANG S K, KANG B H, et al. Schema Mapping Using Hybrid Ripple-Down Rules[C]// *The Australasian Computer Science Conference*. 2015.
- [52] LI J, TANG J, LI Y, et al. RiMOM: A Dynamic Multistrategy Ontology Alignment Framework [J]. *IEEE Transactions on Knowledge & Data Engineering*, 2008, 21(8): 1218-1232.
- [53] MA Z, ZHANG F, YAN L, et al. Fuzzy Knowledge Management for the Semantic Web[J]. *Studies in Fuzziness & Soft Computing*, 2014, 306(S290): 123.
- [54] SHVAIKO P, EUZENAT J. Ontology Matching: State of the Art and Future Challenges[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2013, 25(X): 158-176.
- [55] GAO W, GAO Y, ZHU L. Ranking based ontology learning algorithm for similarity measuring and ontology mapping using representation theory[J]. *Journal of Information and Optimization Sciences*, 2016, 37(2): 303-320.
- [56] GILLANI S, NAEEM M, HABIBULLAH R. Semantic Schema Matching Using DBpedia[J]. *International Journal of Intelligent Systems and Applications*, 2013, 5(4): 72-80.
- [57] FARIA D, PESQUITA C, SANTOS E, et al. The Agreement Maker Light Ontology Matching System[C]// *ODBASE-OTM 2013 Conferences*. 2013: 527-541.
- [58] FARIA D, MARTINS C, NANAVALTY A. AML Results for OAEI 2015[C]// *ISWC International Workshop on Ontology Matching (OM)*, CEUR Workshop Proceedings. 2015.
- [59] WANG P, ZHOU Y, XU B. Matching large ontologies based on reduction anchors[C]// *Proceedings of the 22nd International Joint Conference on Artificial Intelligence(IJCAI 2011)*. Barcelona, Catalonia, Spain. 2011: 140-160.
- [60] WANG W, WANG P. Lily Results for OAEI 2015[C]// *International Workshop on Ontology Matching*. 2015.
- [61] CHEATHAM M, DRAGISIC Z, EUZENAT J. Results of the Ontology Alignment Evaluation Initiative 2015[C]// *Proc. 8th ISWC workshop on ontology matching (OM)*. 2015: 61-100.
- [62] XIANG C C, WEI Z F, ZHAN W D. On Mapping between HowNet and CCD[J]. *Journal of Chinese Information Processing*, 2015, 29(3): 44-51. (in Chinese)
- 向春丞, 魏志方, 詹卫东. HowNet 与 CCD 映射方法研究[J]. *中文信息学报*, 2015, 29(3): 44-51.
- [63] ZHANG L Y, MA Z S, CHEN S X. A Method of Ontology Mapping based on Multiple-Matchers[J]. *Journal of Zhengzhou University*, 2015, 36(3): 106-109. (in Chinese)
- 张凌宇, 马志晟, 陈淑鑫. 一种基于多种类型匹配器的本体映射方法[J]. *郑州大学学报(工学版)*, 2015, 36(3): 106-109.
- [64] LI J, TANG J, LI Y. RiMOM: A dynamic multistrategy ontology alignment framework[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2009, 21(8): 1218-1232.