

本体映射系统的评价体系研究

黄 奇¹, 范佳林², 陆佳莹³, 钱韵洁³

(1. 南京大学国家信息资源管理南京研究基地, 南京 210023; 2. 南京大学工程管理学院, 南京 210093;
3. 南京大学信息管理学院, 南京 210023)

摘 要 本文在辨析了本体映射含义以及概述了本体映射评价研究现状的基础上, 针对目前本体映射评价角度比较单一, 提出了一套本体映射系统的评价体系。为了说明评价体系的合理性, 提出了映射强度的概念, 并对各个层面及其内容指标进行解读介绍。通过对各层面指标进行量化说明, 进一步检验了评价体系的可用性, 并通过对具体映射系统中的应用, 发现了未来的研究方向。

关键词 本体映射; 评价体系; 映射强度; 指标量化

Evaluation System Research of an Ontology Mapping System

Huang Qi¹, Fan Jialin², Lu Jiaying³ and Qian Yunjie³

(1. National Information Resource Management Nanjing Research Base of Nanjing University, Nanjing 210023;
2. School of Management & Engineering in Nanjing University, Nanjing 210093;
3. School of Information Management in Nanjing University, Nanjing 210023)

Abstract: Based on the analysis of the meaning of ontology mapping and the current situation of ontology mapping evaluation, differing from the current relatively singular angle of ontology mapping evaluation, an evaluation system for ontology mapping systems is proposed. In order to explain the rationality of the evaluation system, the concept of mapping intensity is put forward and the interpretation of each level and its content index is introduced. Through the quantification of the indicators at each level, the usability of the evaluation system was tested and the future research direction was found through the application to the specific mapping system.

Key words: ontology mapping; evaluation system; mapping intensity; indicator quantification

1 引 言

近年来, 物联网 (Internet of Things, IoT), 作为信息产业发展的第三次浪潮, 为人们关注的焦点。顾名思义, 物联网就是物物相连的互联网, 它需要各个领域的产品本体之间进行信息交换和资源共享。可是目前, 随着各个领域本体数目的不断增多

以及规模的日益扩大, 不同本体之间产生了大量的信息冗余和重叠, 为物联网的发展造成一定阻碍。为解决这一问题, 相关的学者将精力转移到本体映射的研究上, 使之成为目前信息学界的研究热点。

目前关于本体映射的研究已经较为深入, 研究者们开发出了若干本体映射的系统, 如 RIMOM、PROMPT 以及 IF-MAP 等, 其中一些映射系统已经

收稿日期: 2016-11-26; 修回日期: 2017-04-09

基金项目: 国家社会科学基金重点项目“基于知识组织的产品分类本体研究”(13ATQ005), 教育部人文社会科学研究项目“产品分类的语义分析及其形式化研究”(11YJA870014)。

作者简介: 黄奇, 男, 1961 年生, 博士, 教授, 主要研究方向为网络信息资源的语义化、电子商务; 范佳林, 男, 1993 年生, 在读研究生, 主要研究方向为网络信息资源的语义化、电子商务, E-mail: fan_frankie@163.com; 陆佳莹, 女, 1992 年生, 在读研究生, 主要研究方向为网络信息资源的语义化、电子商务; 钱韵洁, 女, 1994 年生, 在读研究生, 主要研究方向为网络信息资源的语义化、电子商务。

得到较为广泛的应用。学界从两个方面对本体映射系统进行评价,一是从本体映射方法的角度进行评价,二是从本体映射结果的角度进行评价。对于前一方面,只是从不同角度对映射方法进行分类^[1-3]:根据本体定义模型分类(基于类、实例或是基于语法、结构层等)以及根据采用的技术分类(采用统计学方法或是机器学习方法等),这两种分类都是较为简单的总结概述,缺少进一步的比较分析。对于后一方面,目前应用最广泛的评价标准是源自信息检索领域的查准率(P值)和查全率(R值)^[4]进行评价,这种方式只从系统的最终结果进行比较分析,比较面较为单一。因此,目前对于本体映射系统的评价,并没有形成一种较为完整的评价体系。

本文针对本体映射系统评价研究现状,提出一套本体映射系统的评价体系,融合了本体映射的方法和结果两种评价角度,并从更接近本体本身的视角把本体映射的评价工作提升到集词汇层面、结构层面、语义层面和实用性等四个层面为一体的全局体系的高度,并进一步进行量化说明。最后借此评价体系对三种主要的本体映射系统进行相应的分析研究。

2 本体映射

2.1 含义

目前学术界关于本体映射(ontology mapping)的看法主要有狭义和广义两个方面。从狭义角度来看,本体映射仅包括本体匹配(ontology matching),只是建立了源本体和目标本体间的对应关系;从广义角度看,本体映射还应该包含本体对齐(ontology alignment)、本体合并(ontology merging),其不仅建立了映射的对应关系,还可以在此基础上,按照映射的对应关系对本体进行类目、属性和公理等的结构上的调整,从而形成一个新的本体。而本文所讨论的本体映射是狭义上的本体映射,其表示形式如图1所示。

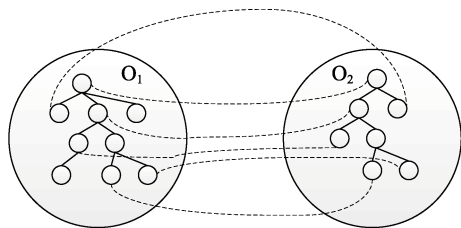


图1 本体映射的示意图

2.2 目的、本质和过程

本体映射是实现不同本体间连接的纽带,是将源本体中实体(类、属性、实例)映射对应到目标本体的过程。根据 Gruber^[5]提出的经典的本体五元组模型,本体映射的类型有类—类、属性—属性、实例—实例、情景和约束等。本体映射的主要目的是实现不同本体间的知识交换,实现信息的重用性和互操作性。其本质问题为相似度的算法问题^[6],即如何确定两个本体间的对应关系。目前关于本体映射的相似度研究已经比较深入,关于相似度特征值,其切入点主要包括基于词汇层面(类、属性、标签说明等)、基于结构层面(结点间位置、距离等)以及语义层面(形式概念分析等),其特征提取的准确性高低将直接影响本体映射的准确性。此外,在映射的过程中,还需要对映射过程中的中间结果不断的迭代修正(阈值、权重值调整等)来增加映射结果的说服力。因此,本体映射的一般过程应该包括特征提取、相似度计算与匹配和循环迭代修正,同时考虑到对映射结果的利用和对原始源本体和目标本体的导入,该过程还应该包括输入和输出,可以用图2进行描述。

3 评价体系的建立与量化

本体映射系统评价体系的建立以本体映射的本质问题为切入点,即如何确定两个本体间的对应关系。由于这一问题主要是依据词汇层面和结构层面的相似,因此把词汇层面和结构层面作为整体评价体系的基础。但词汇层面和结构层面无法进一步揭示本体间深层次的语义关系,因而有必要将语义层面纳入评价体系。另外从使用的角度需考虑语言兼容性、用户体验等多个方面,所以还要兼备映射系统的实用性这一评价层面。综上所述,本文提出的本体映射系统评价体系主要包括了词汇层面的多样性、结构层面的深入性、语义层面的关联性以及映射系统的实用性等四个层面,融合了目前本体映射系统中关于映射方法和映射结果两个方面,并进一步进行拓展,丰富了本体映射系统评价的内涵。本体映射系统评价体系的具体内容以及结构关系如图3所示。

词汇层面的多样性表明映射方法在实体定义的对应比较上是否全面;结构层面的深入性体现映射方法在结构关系层面挖掘的深度情况。这两种层面是整体评价指标的基础,均直接影响着语义层面的关联性,也对映射最终效果的优劣起决定性作用。

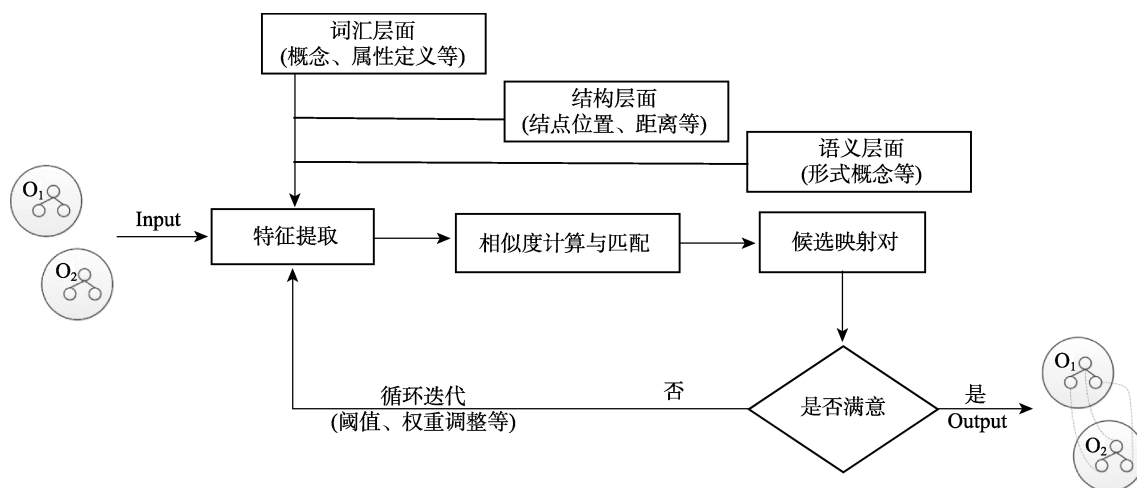


图2 本体映射的一般过程

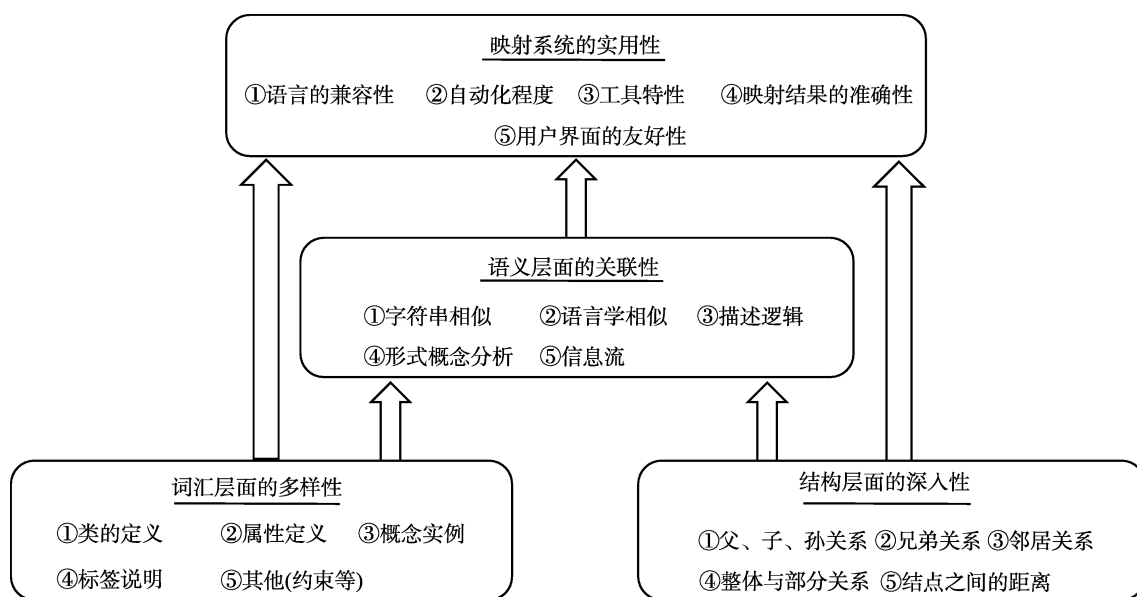


图3 本体映射系统的评价体系

语义层面的关联性是对前两者的延伸和递进，能够弥补词汇层面和结构层面的不足，使得映射更加完美；映射系统的实用性是从实际应用角度出发，是前三个层面的综合性体现。

本文还提出了映射强度概念，旨在将各层面的映射指标与本体映射的全面程度联系起来，既能够增强本体映射系统评价体系中具体指标的说服力，也为下文的指标量化奠定基础。

3.1 映射强度

映射强度（Mapping Intensity, MI），是为了表示两个本体间的映射程度而提出的，它是衡量映射结果的全面性和准确性的综合性概念，是“满意度”的表现，即映射强度越大，表示本体之间的映射程

度越高，映射结果也就越全面准确，越令人满意。映射强度的概念可以用两元组简单进行表示：

$$M_{O_2}^{O_1}(a) = \{s_a, \text{num}(a_i)\}$$

其中， a 表示本体映射系统评价体系中的某个层面， $M_{O_2}^{O_1}(a)$ 表示本体 O_1 和本体 O_2 之间基于 a 层面的映射强度， a 层面里的具体映射指标表示为 a_i ； s_a 表示本体间基于 a 层面的映射研究深度，是 a 层面里的映射指标 a_i 共同作用的结果； $\text{num}(a_i)$ 表示基于 a 层面的映射过程中考虑的映射指标 a_i 的个数。总之， $\text{num}(a_i)$ 和 s_a 越大， $M_{O_2}^{O_1}(a)$ 越大，即考虑 a 层面里的具体映射指标越多，且本体间基于 a 层面的映射研究越深，那么本体 O_1 和本体 O_2 之间基于 a 层面的映射强度也就越高。

3.2 评价体系

3.2.1 词汇层面的多样性

基于词汇层面进行相似度映射是本体映射系统开发的主要思想之一,如 GLUE 系统^[7]等,而且基于词汇层面中的多个方面进行的相似度映射,能够为本体映射提供强大的说服力和评判力度,为映射系统的实用性提供保障,因此词汇层面的多样性指标就是衡量本体映射系统的基础评价指标。本文结合本体构建的基础流程来考虑词汇层面的映射指标组成,所以词汇层面多样性的具体指标内容是从类的定义开始,依次包括属性定义、实例、标签说明以及其他(约束等)方面。随着往后考虑的因素越多,其研究的深度不断的加深,那么词汇层面的映射强度也逐渐变高,如图4所示。

3.2.2 结构层面的深入性

基于结构层进行相似度映射,同样是目前学界进行本体映射的主要思想,在本体映射系统评价体系中词汇层面具有同等地位。结构层面的深入性主要是描述本体映射系统在评判源本体和目标本体之间相似度时对于本体结构的依赖部分,即基于本体结构的哪些特征来判别相似度。与词汇层面相似,结构层面的映射指标也是按照从基础向全面进行建立。因此评价体系中,结构层次的深入性的评价指标主要包括父子孙关系(is-a)、兄弟关系(brother)、邻居关系(neighborhood)、整体和部分的的关系(part-of)以及结点之间的距离关系等。而随着结构层面的映射指标的增多,研究也相对愈加深,因此结构层面映射强度自然也是逐渐变高的过程,如图4所示。

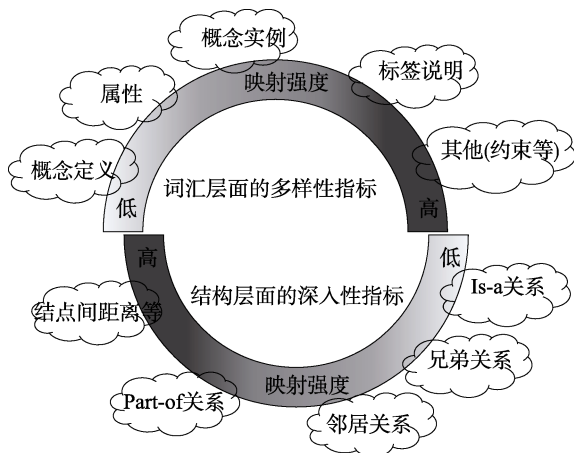


图4 词汇层面和结构层面的映射强度

3.2.3 语义层面的关联性

语义层面的关联性,表示能够准确表述本体间语义关系的能力大小,它是词汇层面和结构层面的递进和补充。从语义层面的不同角度进行本体映射,影响着两个本体间的语义关联程度,也决定着最后本体映射的准确性。目前从语义层面进行本体映射研究的案例也有不少,其主要应用的理论方法主要有字符串相似(String Similarity, SS)、语言学相似(Linguistic Similarity, LS)、形式概念分析(Formal Concept Analysis, FCA)、信息流(Information Flow, IF)以及描述逻辑(Description Logic, DL)等。上述的理论方法具有不同的语义强度^[8],一般来说,某理论方法的语义强度高,表明它挖掘本体间的语义程度越深,那么同等条件下,该层面的本体映射强度也会越高。结合上述方法的语义强度的大小顺序,秉承由浅入深,由简单到复杂的理念,语义层面的映射强度表示形式如图5所示。

3.2.4 映射系统的实用性

映射系统的实用性是从用户使用的角度对本体映射系统进一步进行评价,它是基于前面三个层面的综合性体现。映射系统方法的实用性包括语言的兼容性、自动化程度、工具的特性、映射结果的准确性和用户界面的友好性等。

语言的兼容性表示是否支持不同的本体描述语言;系统工具的自动化程度分为半自动化和自动化两种,不考虑手动情况;系统工具的特性包括系统工具的普适性和拓展性,普适性表示系统工具是否容易获取以及是否容易搭建使用,拓展性则表示是否能够依据实际情况进行修改和二次开发;映射结果的准确性表示本体映射结果的P值、R值的大小;用户界面的友好性旨在能够让非专业的领域学者都能够进行操作。

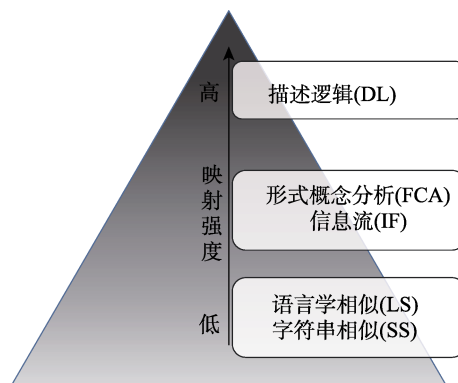


图5 语义层面的映射强度

3.3 指标的量化

为了验证上述评价体系的可使用性,实现对本体映射系统的更深入、直观的评价分析,本文结合了映射强度概念,对上述四种指标进行量化,从而形成一套量化的指标体系,如表1所示。其中 P_i 表示四个映射指标的层面比率, α_{ij} 表示本体映射系统评价体系中各个层面中的具体评价指标, β_{ij} 表示不同的评价指标在所在层面中所占的权重。

3.3.1 指标量化的依据

应用AHP层次分析法,以体系结构介绍和映射强度概念介绍为背景引导进行问卷设计,然后通过调查后的问卷数据的整合和计算,对本体映射系统评价体系中的词汇层面、结构层面、语义层面以及映射系统的实用性等四个层面的比率分别进行量化。之后,再针对每个层面中的具体指标进行进一步的权重量化。

(1) 层面比率 P_i

层面比率衡量的是各个层面在评价体系中的重要程度,重要性越大的层面,其所占的比率越高。层面比率大小的判定主要依据四个层面在评价体系中具有的地位特征,例如,词汇层面和结构层面是

整个评价体系的基础,共同支撑着整体的评价体系,所以两者同等重要,且重要性较大;语义层面的关联性是在前两者的延伸和递进,具有弥补作用;映射系统的实用性是进一步从实用角度出发,是前面三个层面所取得效果的综合性体现。基于这些特征,通过对问卷数据的整合和计算,得到的判断矩阵如图6所示(斜线部分的数值与关于主对角线对称位置的数值互为倒数,故省略,下同)。

(2) 评价指标 α_{ij}

每个层面里面的具体指标分别进行规则量化说明,以数值的大小表示指标的达成程度。比如,在评价体系中的前三个层面里,依据目前学界对各个本体映射系统的研究现况,若本体映射系统在某一层面涉及其中的某个指标,则得1分,否则不得分。而在评价体系中的实用性层面里,则需要利用调查问卷的形式,利用目前本体映射系统的研究现况作为背景说明,对被调查者进行基本引导,让其依据自己的主观感受对每个指标在0~1分范围内进行打分。前三种层面的评价指标分数相对公正客观一些,实用性层面的评价指标分数则相对主观。然后对得到的统计数据筛选、汇总和整合,从而得到最后的评判得分。

表1 本体映射系统评价体系的量化指标模型

映射指标	层面比率 P_i	评价指标 α_{ij}	α_{ij} 评价说明	指标权重 β_{ij}	量化结果 φ_i
词汇层面	$P_1=0.324$	α_{11} : 类的定义	如果涉及此指标,则得1分,否则不得分	0.489	$\varphi_1 = P_1 \sum_{j=1}^5 \alpha_{1j} \beta_{1j}$
		α_{12} : 属性定义		0.272	
		α_{13} : 概念实例		0.127	
		α_{14} : 标签说明		0.064	
		α_{15} : 其他		0.078	
结构层面	$P_2=0.324$	α_{21} : 父子孙关系	如果涉及此指标,则得1分,否则不得分	0.47	$\varphi_2 = P_2 \sum_{j=1}^5 \alpha_{2j} \beta_{2j}$
		α_{22} : 兄弟关系		0.231	
		α_{23} : 邻居关系		0.164	
		α_{24} : 整体部分关系		0.076	
		α_{25} : 结点距离		0.059	
语义层面	$P_3=0.14$	α_{31} : 字符串相似	如果涉及此指标,则得1分,否则不得分	0.109	$\varphi_3 = P_3 \sum_{j=1}^5 \alpha_{3j} \beta_{3j}$
		α_{32} : 语言学相似		0.109	
		α_{33} : 信息流		0.207	
		α_{34} : 形式概念分析		0.207	
		α_{35} : 描述逻辑		0.368	
实用性	$P_4=0.212$	α_{41} : 语言兼容性	自动化程度依据数值进行打分,未有数值的项,依照实际情况在0~1内打分	0.36	$\varphi_4 = P_4 \sum_{j=1}^5 \alpha_{4j} \beta_{4j}$
		α_{42} : 自动化程度: 半自动化(0.6) 自动化(1)		0.064	
		α_{43} : 工具特性		0.152	
		α_{44} : 结果准确性		0.36	
		α_{45} : 界面友好性		0.064	

	P_1	P_2	P_3	P_4	W
P_1	1	1	2.2	1.6	0.324
P_2		1	2.2	1.6	0.324
P_3			1	0.6	0.14
P_4				1	0.212

($\lambda_{\max}=4.005$, C.I.=0.002, R.I.=0.900, C.R.=0.002<0.1)

图6 层面比率的判断矩阵

(3) 指标权重 β_{ij}

指标权重的衡量标准主要是结合上文中的映射强度的概念以及目前对于各个层面指标的研究情况。例如词汇层面和结构层面的指标权重按照图4中映射强度的顺序进行衡量,该层面映射强度中靠前的部分,指标权重相对较大,因为越是基础的部分,研究相对越深入,被认可的程度越高,所以设定的指标权重相对较大。语义层面的指标权重也是结合该层面映射强度进行划分的,但与前两个层面不同的是,由于语义层面指标间并没有依赖延续的关系,而存在语义强度的差异,所以图5映射强度的顺序中偏后的指标,其指标权重反而相对较大。最后实用性层面的指标权重,主要从本体映射系统设计目的角度出发,对该层面里的具体指标进行相对必要性的顺序划分,越贴近设计初衷的指标,必要性越高,相应的指标权重越大。将上述指标权重的叙述作为背景引导进行问卷调查,最后得到各个层面的判断矩阵如图7~图10所示。

	β_{11}	β_{12}	β_{13}	β_{14}	β_{15}	W
β_{11}	1	3.2	4.6	6	6.2	0.489
β_{12}		1	3.6	5	5.2	0.272
β_{13}			1	3	3.6	0.127
β_{14}				1	1.8	0.064
β_{15}					1	0.048

($\lambda_{\max}=5.71$, C.I.=0.068, R.I.=1.120, C.R.=0.061<0.1)

图7 词汇层面的判断矩阵

	β_{21}	β_{22}	β_{23}	β_{24}	β_{25}	W
β_{21}	1	3	3.8	5	5	0.47
β_{22}		1	1.6	3.8	4	0.231
β_{23}			1	3	3	0.164
β_{24}				1	1.8	0.076
β_{25}					1	0.059

($\lambda_{\max}=5.159$, C.I.=0.04, R.I.=1.120, C.R.=0.036<0.1)

图8 结构层面的判断矩阵

	β_{31}	β_{32}	β_{33}	β_{34}	β_{35}	W
β_{31}	1	1	1/2	1/2	1/3	0.109
β_{32}		1	1/2	1/2	1/3	0.109
β_{33}			1	1	1/2	0.207
β_{34}				1	1/2	0.207
β_{35}					1	0.368

($\lambda_{\max}=5.013$, C.I.=0.003, R.I.=1.120, C.R.=0.003<0.1)

图9 语义层面的判断矩阵

	β_{41}	β_{42}	β_{43}	β_{44}	β_{45}	W
β_{41}	1	5	3	1	5	0.36
β_{42}		1	1/3	1/5	1	0.064
β_{43}			1	1/3	3	0.152
β_{44}				1	5	0.36
β_{45}					1	0.064

($\lambda_{\max}=5.055$, C.I.=0.014, R.I.=1.120, C.R.=0.013<0.1)

图10 实用性的判断矩阵

3.3.2 混合策略说明

所谓混合策略,指的是两个本体不是基于词汇层面和结构层面中的某单个层面进行本体映射,而是同时考虑两个层面,并依据结果择优选择对应层面,而且目前的很多映射系统都采取这种策略。结合本文的本体映射系统的评价体系,可以发现词汇层面和结构层面同处于基础的等价地位。因此,在计算混合策略的量化指标时,我们只需要计算词汇层面和结构层面中的最大值即可。

3.3.3 指标量化的条件约束

$$(1) \sum_{i=1}^4 P_i = 1$$

$$(2) \sum_{j=1}^5 \beta_{ij} = 1 (i=1,2,3,4)$$

$$(3) \max\{\alpha_{ij}\} = 1 (i=1,2,3,4 \quad j=1,2,3,4,5 \quad \alpha_{ij} \geq 0)$$

$$(4) \text{总得分为:}$$

$$\varphi_{\text{总}} = \max\{\varphi_1, \varphi_2\} + \sum_{i=3}^4 \varphi_i =$$

$$\max\left\{P_1 \sum_{j=1}^5 \alpha_{1j} \beta_{1j}, P_2 \sum_{j=1}^5 \alpha_{2j} \beta_{2j}\right\} + \sum_{i=3}^4 P_i \sum_{j=1}^5 \alpha_{ij} \beta_{ij}$$

4 评价体系的应用与分析

4.1 系统工具的量化指标评分

经过多年不断发展完善,目前在本体映射领域中相对成熟的映射系统工具有60多种^[9],本研究主

要介绍较有特色的三种映射系统: PROMPT、IF-MAP 和 RIMOM, 并从评价体系中的四个层面对每个系统进行量化评分。

4.1.1 PROMPT

PROMPT 系统是斯坦福大学医学信息组开发的、基于字符串匹配和锚点遍历相结合的映射系统^[10], 它是一种半自动的系统, 使映射自动化程度和准确性上在一定程度上得到了平衡。

(1) 词汇层面的多样性: PROMPT 系统中支持 Lexical Matching, 其中涉及类、属性定义、概念实例以及标签说明等, 且在系统给出候选映射对后, 可以进一步对类-类(class-class)、属性-属性(slot-slot)、实例-实例(instance-instance)进行人工增添删改, 所以这一层面的得分为满分 $0.324 \times 1 = 0.324$ 。

(2) 结构层面的深入性: 在 PROMPT 系统中, 通过系统自带的插件, 可以推断其中的父子孙关系、兄弟关系以及邻居关系, 无法在结构层面的映射过程中涉及整体和部分关系以及结点距离, 所以其得分为 $0.324 \times (0.47 + 0.231 + 0.164) = 0.280$, 由此可知 PROMPT 结构层面的得分小于词汇层面的得分, 所以依旧按照词汇层面的得分进行评价。

(3) 语义层面的关联性: PROMPT 系统中对于本体映射的语义层面, 主要是采用字符串相似以及语言学相似, 其他方面并不涉及, 所以这一层面的得分为 $0.14 \times 0.109 \times 2 = 0.031$ 。

(4) 映射系统的实用性: PROMPT 能够支持多种本体语言, 是半自动化, 而且工具易获得、易搭建, 支持二次开发, 普适性和拓展性都很好。本体映射结果的准确性较高^[10], 界面的友好程度在这一领域的系统中可以算是较高水平。所以具体得分为 $0.212 \times (0.36 + 0.6 \times 0.064 + 0.152 + 0.8 \times 0.36 + 0.8 \times 0.064) = 0.189$ 。

综上所述, PROMPT 本体映射系统的量化指标的总分为 0.544。

4.1.2 IF-MAP

IF-MAP 系统顾名思义, 它是一个基于语义信息流的数学理论的自动化本体映射系统^[11], 其使用了一个参考本体作为公共参考本体来对两个输入的本体进行匹配。它建立本体映射的主要步骤为: 本体获取、转化和信息态射(info-morphism)。

(1) 词汇层面的多样性: IF-MAP 本体映射系统在词汇层面主要涉及类的定义、属性定义以及概念实例这三个方面, 没有涉及标签说明及其以外的其

他方面, 故 IF-MAP 在这一层次的得分为 $0.324 \times (0.489 + 0.272 + 0.127) = 0.288$ 。

(2) 结构层面的深入性: IF-MAP 本体映射在结构层面上涉及了类之间的父子孙关系、兄弟关系和邻居关系, 在其他方面没有涉及, 所以在结构层面的得分为 $0.324 \times (0.47 + 0.231 + 0.164) = 0.28$ 。故 IF-MAP 结构层面的得分小于词汇层面的得分, 以词汇层面的得分为主。

(3) 语义层面的关联性: 除了应用到信息流的数学理论方法, IF-MAP 系统还涉及了较为基础的字符串相似和语言学相似。所以这一层面的得分为 $0.14 \times (0.109 + 0.109 + 0.207) = 0.06$ 。

(4) 映射系统的实用性: IF-MAP 系统支持多种本体语言, 为半自动化程度的系统, 但是目前工具尚未成熟, 并无合适的支持环境^[12], 更不用提二次开发。本体的映射结果较为准确^[11], 用户界面的友好性由于工具特性较差, 无法获取, 所以这一指标可以忽略不计, 记为无。这一层面的得分为 $0.212 \times (0.36 \times 1 + 0.064 + 0.36 \times 1) = 0.166$ 。

综上所述, IF-MAP 本体映射系统的量化指标的总分为 0.514。

4.1.3 RIMOM

RIMOM 是由清华大学开发出来的本体自动化映射工具, 是基于贝叶斯决策理论提出最小风险的本体映射模型^[13]。RIMOM 将映射发现问题转换成风险最小化问题, 同时可以给出不同策略的本体映射方法。

(1) 词汇层面的多样性: RIMOM 的映射策略中包括基于类、属性定义、概念实例以及标签说明等, 同时, 它还包括基于约束的策略, 所以 RIMOM 在词汇层面的多样性程度很高, 所以其在词汇层面的得分和 PROMPT 一样, 为 0.324。

(2) 结构层面的深入性: RIMOM 中有一种策略为结构上下文映射策略, 其主要的思想就是考虑父子类、兄弟关系、邻居关系以及属性等。考虑到 part-of 也为属性, 所以这一层面 RIMOM 映射系统的得分为 $0.324 \times (0.47 + 0.231 + 0.164 + 0.076) = 0.305$ 。因此按照词汇层面的得分进行计算。

(3) 语义层面的关联性: RIMOM 结合了概念词典, 其中涉及了字符串相似和语言学相似, 在形式概念分析、信息流以及描述逻辑方面均不涉及, 语义层面和 PROMPT 相似, 其关联性比较弱一些。因此在这一层面上, RIMOM 映射系统的得分为 $0.14 \times 0.109 \times 2 = 0.031$ 。

(4) 映射系统的实用性: RIMOM 支持多种本体语言, 它是自动化系统。在系统工具特性方面, 从普适性和拓展性的角度都比 PROMPT 较弱一些, 比 IF-MAP 强一些。映射结果的准确性方面处于较高水平^[13], 由于是系统模型, 所以在用户界面的友好性我们就暂时不考虑。所以这一层面的得分为: $0.212 \times (0.36 \times 1 + 0.064 + 0.5 \times 0.152 + 0.8 \times 0.36) = 0.167$ 。

综上所述, RIMOM 本体映射系统的量化指标的总分为 0.522。

4.2 评价体系的应用情况分析

下文的表 2~表 5 概括了三种映射系统在此评价体系中涵盖的具体方面, 这有助于我们从整体的角度对三种系统工具进行比较和评价。此外, 以这三种具体的映射系统为镜, 可以折射出目前的主流映射系统的主要表现, 为进一步的研究提供方向。

表 2 词汇层面的多样性汇总

	类的定义	属性定义	概念实例	标签说明	其他
PROMPT	√	√	√	√	人工增添
IF-MAP	√	√	√		
RIMOM	√	√	√	√	约束

表 3 结构层面的深入性汇总

	is-a 关系	兄弟关系	邻居关系	整体部分关系	结点距离
PROMPT	√	√	√		
IF-MAP	√	√	√		
RIMOM	√	√	√	√	

表 4 语义层面的关联性汇总

	字符串相似	语言学相似	信息流	形式概念分析	描述逻辑
PROMPT	√	√			
IF-MAP	√	√	√		
RIMOM	√	√			

表 5 映射系统的实用性汇总

	语言兼容性	自动化程度	工具特性	结果准确性	界面友好性
PROMPT	多种	半自动化	普适性和拓展性都很高	较高	很好
IF-MAP	多种	自动化	较低	较高	无
RIMOM	多种	自动化	较低	很高	无

(1) 从上述的比较可以看出, 在指标评价的整体得分上, PROMPT 相对于其他两个映射系统具有较大的优势, 这也符合目前 PROMPT 系统比其他两个系统应用更普遍的现状。

(2) 在同为体系基础的词汇层面和结构层面上, 词汇层面可能比结构层面研究的更深入一些, 如 PROMPT 和 RIMOM 均得满分。而在结构层面上, 父子孙关系、兄弟关系和邻居关系的研究相对较多, 而在 part-of 等属性关系方面以及结点之间的距离方面的研究, 还有待于进一步的加强和深化。

(3) 语义层面上, PROMPT、RIMOM 以及 IF-MAP 本体映射系统都集中在字符串相似和语言学相似等基础的理论方法上, 其居于映射强度顺序前段位置, 语义关系挖掘较浅。信息流理论仅在 IF-MAP 中涉及, 而描述逻辑与形式概念分析理论方法目前也分别仅在 RacerPro^[14]和 FCA-Merge 系统^[15]中涉及, 所以后三种理论方法虽然居于映射强度顺序后段位置, 语义强度较高, 但在映射系统中的应用情况还需要学者们的继续探索完善。

(4) 关于映射系统的实用性层面, 本体映射系统的自动化程度已普遍较高, 但是很多系统仍停留在系统模型层面, 如 IF-MAP 和 RIMOM, 如何将之转变为完整的现实系统并进行推广应用也是未来继续研究的方向。在这一方面, PROMPT 的系统较为完善, 且目前已经得到相对广泛的应用。

综上所述可以发现, 如何进一步挖掘基于结构层面的本体映射, 以及如何将信息流、形式概念分析和逻辑描述等语义强度较高的理论方法应用到本体映射系统中值得学者们进一步探索, 是未来的研究方向。

5 结束语

对于本体映射系统的评价, 本文建立了统一的评价体系, 突破了目前本体映射系统间自我评价标准差异化的现况, 有助于用户从公共的视角对本体映射系统进行选择和比较。此外, 本文将传统的基于映射方法或映射结果的单一评价指标提升为基于集词汇层、结构层、语义层和实用性等相互联系的四个层面为一体的评价体系, 丰富了各个层面里的内容, 并阐明了各个层面间的关系, 希冀能够对拓展目前本体映射系统评价的研究思路有所贡献。

参 考 文 献

- [1] 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, 33.
- [2] Kalfoglou Y, Schorlemmer M. Ontology mapping: the state of the art[J]. The Knowledge Engineering Review, 2003, 18(1): 1-31.
- [3] Perez A G. Deliverable 1.3: A survey on ontology tools[EB/OL]. [2002-05-31]. <http://www.aifb.kit.edu/web/deliverable714>.
- [4] 邓汉成, 王敏芳, 王琰. 查全率与查准率之间关系的理论研究[J]. 情报学报, 2000, 19(4): 359-362.
- [5] Gruber T R. A translation approach to portable ontology specifications[J]. Knowledge Acquisition, 1993, 5(5): 199-220.
- [6] 蔡丽宏, 马静. 基于综合方法的一种本体映射实验研究[J]. 情报学报, 2010, 29(5): 820-825.
- [7] Doan A H, Madhavan J, Domingos P, et al. Learning to map between ontologies on the semantic web[C]// Proceedings of the 11th International Conference on World Wide Web. New York: ACM Press, 2002: 662-673.
- [8] Kalfoglou Y, Hu B, Reynolds D, et al. Semantic integration technologies survey[R]. Southampton: University of Southampton, 2005.
- [9] Otero-Cerdeira L, Rodríguez-Martínez F J, Gómez-Rodríguez A. Ontology matching: A literature review[J]. Expert Systems with Applications, 2015, 42(2): 949-971.
- [10] Noy N F, Musen M A. The PROMPT suite: interactive tools for ontology merging and mapping[J]. International Journal of Human-Computer Studies, 2003, 59(6): 983-1024.
- [11] Kalfoglou Y, Schorlemmer M. IF-MAP: an ontology-mapping method based on information-flow theory[J]. Journal on Data Semantics, 2003, 1(1): 98-127.
- [12] 米杨, 曹锦丹. 基于 PROMPT 的本体映射实例分析[J]. 情报学报, 2010, 29(6): 987-991.
- [13] 唐杰, 梁邦勇, 李涓子, 等. 语义 Web 中的本体自动映射[J]. 计算机学报, 2006, 29(11): 1956-1976.
- [14] 沈国华, 黄志球, 朱小栋, 等. 一种应用描述逻辑推理实现本体映射的方法[J]. 小型微型计算机系统, 2008, 29(11): 1979-1985.
- [15] Stumme G, Maedche A. FCA-MERGE: bottom-up merging of ontologies[C]// Proceedings of the 17th International Joint Conference on Artificial Intelligence. San Francisco: Morgan Kaufmann Publishers, 2010: 225-230.

(责任编辑 马 兰)