

本体应用综述*

A Survey on Ontology Application Research

周相兵

ZHOU Xiang-bing

(阿坝师范高等专科学校计算机科学系,四川成都 611741)

(Department of Computer Science, Aba Teachers College, Chengdu, Sichuan, 611741, China)

摘要:从本体应用基础、本体应用的相似度计算方法、本体应用领域及存在问题 3 个角度对本体的应用进行综述,指出本体将在智能化识别推理、产品设计和材料搭配、信息和知识管理、云计算和物联网等 5 个方面得到更广泛、更深层次的应用。

关键词:本体 应用 相似度 展望

中图分类号:TP311 **文献标识码:**A **文章编号:**1002-7378(2010)04-0523-05

Abstract: In the paper, ontology application from three different points including the basis of ontology application, similarity of ontology, application fields of ontology and some problems of ontology are summerized. The application ontology in five different points such as recognition and reasoning, product design and materials configuration, information and knowledge management, cloud computing and internet of things is illustrated.

Key words: ontology, application, similarity, prospect

本体(Ontology)是从哲学概念中引入到计算机科学领域的,被 Studer 等人定义为共享概念模型明确的形式化规范说明^[1],目前已经在人工智能、语义 Web、软件工程、生物医学信息学、图书馆学以及信息架构等领域得到了广泛应用。其中基于本体的语义 Web 研究是当前研究热点之一,而语义 Web 是由 Tim Berners-Lee^[2]提出的一种实现数据和信息自动处理的方式。要实现语义 Web 的功能,需要提供一种计算机能够理解的、结构化的语义描述机制,以及一系列的推理规则以实现自动化推理^[3,4]。因此,在语义 Web 体系结构中主要由可扩展标记语言(XML)、资源描述框架(RDF(S))和本体来实现信息自动处理及推理,但是 XML 只有语法性而无语义,只能做为数据和信息的载体和规则约束;RDF 可以实现语义描述,由 RDF Data Model、RDF

Schema 和 RDF Syntax 部分组成;本体用来实现语义推理和实现 RDF 语义描述。目前针对 web 服务描述本体的语言有 OWL-S(Ontology Web Language for Services)和 WSMO(Web Service Modeling Ontology)等。其中 OWL-S 是对 RDF、RDF Schema 的进一步扩展,它不但提供基于 XML 的语法级的检测,还提供了语义级的推理,在分布式集成、信息复杂处理、人机对话等领域得到了广泛的应用。本文从本体应用基础、本体应用相似度计算方法和本体应用领域及存在问题 3 个角度对本体的应用进行综述,并对本体更广泛、更深层次的应用进行展望。

1 本体应用基础概述

1.1 本体应用定义

计算机里的本体是以概念化、语义、明确化术语来定义描述的,它由概念、属性、关系、功能、实例、公理以及事件组成。通常由 $O < C, A^C, R, A^R, H^C, X >$ 来描述本体^[5],其中 C 表示概念集, A^C 表示每个概念的属性集, R 表示关系集, A^R 表示每个关系的属性集, H^C 表示概念层次, X 表示对本体的约束公

收稿日期:2010-04-19

修回日期:2010-10-15

作者简介:周相兵(1980-),男,讲师,主要从事语义智能与服务计算研究。

*四川省教育厅自然科学基金项目(09ZC002),阿坝师专校级重点项目(ASA09-15)资助。

理集;本体应用描述为 $OA(O, OL, OR, OT, OE)$, O, OL, OR, OT, OE 分别表示本体定义、本体语言(如 OWL-S 和 WSMO)、本体推理方法、本体编辑工具和本体应用效率评估(如查准率和查全率),其中 $O = \{o_1, o_2, \dots, o_n\}$ 。

1.2 本体描述的数学基础

在应用本体时,描述逻辑(DL)是本体的基础,它是基于概念(类)和角色(属性)的知识表示形式,是构建在框架知识表示方法和语义 Web 的基础上。DL 是对概念化知识进行表示和推理的逻辑形式,可以看做是框架、语义 Web 和一阶逻辑的有机结合。使用特定的 DL 提供的概念和角色构造函数来建立原子概念(一元谓词)和原子角色(二元谓词)的表示式,用语义 Web 和框架等来定义 DL。同时,DL 是阶谓词逻辑的可判定子集,而 DL 系统能够提供可判定的推理服务,在 DL 中引入一阶逻辑不仅解决了语义问题,而且得到了推理机制;这样通过提供清晰的模型论语义,DL 就能够处理结构化概念的表示和推理。因此,语义 Web 本体语言 OIL, DAML+OIL 和 OWL 都建立在描述逻辑基础上,并且 Horrocks I^[6] 证明了 OIL 与描述逻辑 SHIQ 等价、DAML+OIL 与描述逻辑 SHOIQ(D) 等价以及 OWL Lite 与描述逻辑 SHIF(D) 等价、OWL DL 与描述逻辑 SHOIN(D) 等价。

1.3 本体分类

本体建模分类可以分为领域本体(DO)和上层本体(UO)。目前,存在着几种现成可用的标准化上层本体,包括都柏林核心、通用形式化本体、OpenCyc、ResearchCyc、推荐上层合并本体以及 DOLCE。另外 Gellish 本体是一个关于上层本体与领域本体间组合的例子^[7]。在具体应用本体建模时,就需要相关的本体开发工具来实现本体编制,即本体语言。这些语言包括普通逻辑(Common logic)^[8]、基于一阶谓词演算的本体语言(Cyc)^[9]、Gellish^[10]、IDEF5^[11]、知识交换格式(KIF)^[12]、规则交换格式(RIF)与 F-逻辑(F-Logic)^[13]、OWL(Web Ontology Language)^[14] 等。而常用的本体编辑平台有 CO-ODE 本体浏览器(本体服务器)^[15]、Protégé^[16]、Neon 工具箱^[17] 等。本体推理机一般有 Pellet、Racer、FaCT++、JENA 等。

1.4 基于本体的语义 Web 服务描述

与本体紧密相关的应用就是语义 Web,它是近年 WWW 兴起的新技术,因此产生了语义 Web 与 Web 服务相结合的语义 Web 服务,而语义 Web 服

务的目标是在 Web 服务的描述中注入足够的语义信息,使 Web 服务成为计算机可以理解的实体,从而支持 Web 服务的自动发现、自动组合和自动执行等^[3],最终达到多样化的分布式计算。语义 Web 服务的直接表现就是将语义 Web 的特征置入到 Web 服务中去,实现 Web 服务各项工作的自动化。目前,基于本体的 OWL-S 和 WSMO 都对 Web 服务语义描述提供了方法。在 OWL-S 中,从 ServiceProfile、ProcessModel 和 ServiceGrounding 3 个方面对 Web 服务进行刻画,其中,ServiceProfile 类似于服务的黄页,描述服务的功能及相关属性;ProcessModel 对服务的过程模型进行刻画,描述服务是如何工作的;ServiceGrounding 将过程模型与通信协议及消息格式等联系起来,描述如何访问一个服务。在 WSMO 中,从 Ontologies、Web services、Goals 和 Mediators 4 个角度进行刻画语义描述,其中 Mediators 是贯穿整个语义 Web 服务编排和排列的核心;Mediators 是从 Web Compliance、Ontology - Based、Strict Decoupling、Ontological Role Separation、Description versus Implementation、Execution Semantics 和 Service versus Web service 方面进行设计的。OWL-S、WSMO 与 WSDL 的对比如表 1 所示^[4]。在服务消费与服务提供间查询 UDDI 服务信息,以及通过 WSDL 抽取具体的服务功能时,一般都是通过关键词方法查询和匹配较合适的服务,因此文献[3]中引入动态描述逻辑方法来提高服务发现的查准率和查全率。

表 1 OWL-S、WSMO 与 WSDL 的异同对比表^[4]

服务处理项	OWL-S	WSMO	WSDL
服务发现	Profile	目标和 Web 服务 (capability)	UDDI API
服务消费与聚集	过程(流程)模型	服务接口(编排+排列)	BPLAWS/WS-CDL
调用方法	Grounding + WSDL/SOAP	Grounding (WS-DL/SOAP, 基于本体)	SOAP
异构处理		中介器	

2 本体应用相似度的计算方法

在应用本体去完成具体业务逻辑时,通常采用相似度来度量两个本体概念、实例等的结合情况,一般情况以结构相似和语言相似进行量化。本体相似度计算由 $Sim(O_i, O_j, OA(app))$ 描述,其分别表示本体的相似度计算方法和计算方法的选择。常用的相似度计算方法主要有以下 6 种。

(1) 具有路径长度和概念层次深度的计算方法^[18]:

$$\text{Sim}(c_1, c_2) = \begin{cases} e^{\alpha l} \times \frac{e^{\beta h} - e^{-\beta h}}{e^{\beta h} + e^{-\beta h}}, & c_1 \neq c_2; \\ 1, & \text{other,} \end{cases}$$

其中, $\alpha \geq 0, \beta > 0$ 为系数, $l = \text{dist}(c_1, c_2)$ 为两概念间的最小距离, $h = \text{length}(\text{root}, \text{lcs})$ 为 c_1, c_2 的最小公共你概念的深度。这种计算方法通常应用于具有层次结构的本体建模中。

(2) 余弦相似计算机方法^[19]:

$$\text{Sim}(i, j) = \frac{\sum_{c \in I_{ij}} (R_{i,c} - \bar{R}_i)^2 (R_{j,c} - \bar{R}_j)^2}{\sqrt{\sum_{c \in I_i} (R_{i,c} - \bar{R}_i)^2} \sqrt{\sum_{c \in I_j} (R_{j,c} - \bar{R}_j)^2}},$$

其中, i, j 分别表示两个概念, I_{ij} 表示 i, j 共同评分集, I_i 和 I_j 分别表示 i, j 单独评分集, $R_{i,c}$ 表示 i 对概念 C 的评分, $R_{j,c}$ 表示 j 对概念 C 的评分。 \bar{R}_i, \bar{R}_j 分别表示 i, j 对概念平均评分。这种计算方法通常应用于选择性的非层次结构的本体建模中。

(3) 语义距离相似计算方法^[19]:

$$\text{Sim}(c_i, c_j) = \sum_{k=1}^n w_{sk} + \frac{N_{c_i} + N_{c_j}}{N_{c_i} + N_{c_j} + 2N_{LCA}} \epsilon,$$

其中 $\sum_{k=1}^n w_{sk}$ 表示 c_i, c_j 的最短路径中各边的加权距离之和, N_{c_i} 和 N_{c_j} 分别表示 c_i, c_j 到最低共同祖先的加权距离, N_{LCA} 表示最低共同祖先到根节点的加权距离, ϵ 为一个常数。这种计算方法通常用于本体小范围内的树形匹配计算中。

(4) 相似度的层次计算方法^[20]:

$$\begin{aligned} \delta: O \times O &\rightarrow R, \\ \forall e_1, e_2 \in O, \delta(e_1, e_2) &= \\ \min_{c \in O} [\delta(e_1, c), \delta(e_2, c)], & \\ H = \langle O, \leq \rangle, & \end{aligned}$$

其中 $\delta(e, c)$ 表示在两个概念 C 间的许多中间边。这种计算方法通常用于图结构的本体建模中。

(5) 相似度的层次计算方法二^[21]:

$$\begin{aligned} \delta: O \times O &\rightarrow R, \\ \delta(c_1, c_2) &= \frac{UC(c_1, H) \cap UC(c_2, H)}{UC(c_1, H) \cup UC(c_2, H)}, \\ H = \langle O, \leq \rangle, & \end{aligned}$$

其中 $UC(c, H) = \{c_2 \in H; c_1 \leq c_2\}$ 。这种计算方法通常用于具有简单层次结构的本体建模中。

(6) 距离相似计算方法^[22]:

$$\begin{aligned} \text{Sim}(c_1, c_2) &= 1 - \text{dist}(c_1, c_2), \\ \text{dist}(c_1, c_2) &= \text{dist}(c_1, \text{ccp}) + \text{dist}(c_2, \text{ccp}), \\ \text{dist}(c, \text{cpp}) &= \text{miles}(\text{ccp}) - \text{miles}(c), \\ \text{miles}(n) &= \frac{1}{k^{l(n)+1}}, \end{aligned}$$

其中 $l(n)$ 是根到结点的最长距离, k 是一个预定义因子。这种计算方法通常用于可控、简单的本体建模中。

除了上述常见的相似度计算方法外,其它相似度计算方法大多都是采用图结构和树结构进行建模,并通过上述 6 种计算公式变换设计出不同需求的相似度计算方法^[23]。本体应用的相似度计算是建立概念间关系的必要步骤,而怎样建立或选择合适的相似计算方法是需要根据具体应用事物的特征进行分析,但是这大大增加了选择相似度的难度。通常情况下,相似度高的事物(概念),它们的关系才紧密,才容易完成匹配,实现概念间的关系才容易实现逻辑推理。

3 本体应用领域及存在问题分析

3.1 本体应用领域

通过 wikipedia 查询可知,现在已公布的本体库有 8 个,本体例子有 34 个,主要以生物学与植物学相关的本体例子最多,其次是一些基础性和词汇性的本体例子^[7]。若根据本体的独特的性质研究出一种基于本体的语义搜索引擎,将全世界已公布的本体聚集起来,就可以有效实现相关业务功能自动识别、自动推理和自动组合等,这样就会大大提高互联网使用效率^[24]。同时,基于本体的语义检索也是本体应用的具体表现^[25],本体在电子政务^[26]、知识管理^[27]、企业知识集成^[28]、数字博物馆^[29]、异构信息抽取^[30]、产品概念设计^[31]等领域也得到了很好的应用。

3.2 本体应用存在的问题分析

基于本体的应用,在模式识别、人工智能、精密设计、信息精确识别、信息融合、信息管理、逻辑推理等领域具有巨大的推动和促进作用。这些应用主要通过本体概念化描述来模拟事物的特征,使存在的事物具备关系,让它们在物与物间建立可推理的逻辑联系,因此本体适合现今信息科学领域的需求,同时也满足信息科学领域的具体应用。但是在具体应用过程中,因本体形式描述过于抽象,具体表达方式过于模糊,对事物特征描述难以定性,对庞大的信息实体进行概念化难度大,对概念定性化依赖度大等

因素的影响,给本体在信息科学领域广泛、深层次的应用带来了挑战。而本体应用的理论基础是由描述逻辑确定的,它描述了事物间概念化的推理机制,并用数学的方式进行证明推理的可行性、正确性,这对本体应用的可扩展性提供了保障;但是根据事物不同特征出现了多种变形的描述逻辑语言,这就增加了本体描述的多样性,导致了本体应用的难度。特别在用具体描述语言进行描述时,往往需要更深层次的考虑描述语言的兼容性和可用性,如 OWL-S 和 WSMO 都可以描述基于本体的语义 Web 服务,但它们采用的描述方式、方法、原理是不一样的,当然最后应用的效果也是存在差异性。因此在基于本体的具体应用中,首先需要分析事物的本质、特征,并选择合适的描述语言、推理机制;其次根据本体在信息科学里的定义进行抽象具体事物,建立事物的概念,并根据所选择的推理机制进行推理分析;最后完成事物的本体建模并进行应用。

4 展望

本文主要从本体应用基础、本体应用的相似度计算方法和本体的应用领域 3 个角度对本体的应用进行综述。本体具有逻辑识别、自动推理等优点,在很多领域均得到了很好的应用,作者认为本体还将在以下几个方面得到更广泛、更深层次的应用。

(1)在语义 Web 服务中的智能化识别推理的应用。在语义 Web 中,主要起语义作用,使整个 Web 具有语义,形成可操作的智能网、语义网,表现在多 Web 聚类、融合、联接等;在语义 Web 服务中,主要向 WSDL 中注入语义,使 Web 服务概念化,让其能自动识别推理,以达到整个需求过程自动化、精确化。但是目前本体在语义 Web 服务应用领域存在的问题主要表现为:①查准率和查全率难满足用户需求;②相似度计算结果难满足语义匹配的要求;③很多研究成果停留在实验分析上。因此,这方面的应用仍将是今后研究、应用的热点和重点。

(2)产品设计、材料搭配等相关领域的应用。由于本体具备独特的性质,因此在产品精确设计、材料精确搭配等相关领域将得到广泛应用,主要表现在原材料自动搭配、配制,让其满足事物本身的特征。如将多种金属炼制成一种新的金属材料,由于每一种金属所含的金属元素种类不一样,含量不一样,自然就会影响金属的性能,这时就需要按新金属的特征自动搭配一种满足需求的配制方法。

(3)信息、知识管理等领域的应用。信息、知识

管理主要表现在集成、分类、聚类等方面,应用领域有电子商务、图书情报、生物信息学、主题图知识管理^[32]、情感感知^[33]、智能检索^[34]等。

(4)基于云计算的 SaaS(Software as a Service) 组装。云计算是一种新型的、优越的分布式计算模式^[35];而 SaaS 是云计算中的一种具体应用,它具有无维护、无升级、低成本等优势,通常情况下,它是通过 Web 服务接口和 REST(Representational State Transfer)实现 SaaS 组装,因此要有效实现 SaaS 组装建议最好使用基于本体的语义来自动识别 Web 服务接口和 REST,最终满足用户需求。

(5)在物联网中的应用。早期的物联网是依托 RFID(Radio Frequency Identification)技术的物流网络,而现今是指在物理世界的实体中部署具有一定感知能力、计算能力和执行能力的嵌入式芯片和软件,使之成为“智能物体”,并且通过网络设施实现信息传输、协同和处理,从而实现物与物、物与人之间的互联。然而要实现信息协同处理,完成物与物、物与人之间的交互,本体应用将是最好方法之一。

参考文献:

- [1] Gruber T R. A translation approach to portable ontology specifications[J]. Knowledge Acquisition, 1993,5: 199-220.
- [2] Berners-Lee T, Handler J, Lassila O. The semantic web [J]. Scientific American, 2001(184):34-43.
- [3] 史忠植,常亮. 基于动态描述逻辑的语义 Web 服务推理[J]. 计算机学报,2008, 31(9):1599-1611.
- [4] 周相兵,杨兴江,向昌成. 面向本体的 SOA 构件定位机制与算法研究[J]. 系统工程与电子技术,2009,31(6): 1473-1479.
- [5] Alexander Maedche. Ontology learning for the semantic web[M]. Boston/Dordrecht/London:Kluwer Academic Publishers,2002:11-30.
- [6] Horrocks I, Patel Schneider PF, Harmelen FV. From SHIQ and RDF to OWL: the making of a Web ontology language[J]. Journal of Web Semantics, 2003, 1(1):7-26.
- [7] wikipedia. Ontology[EB/OL]. [2010-03-08]. <http://zh.wikipedia.org>.
- [8] Harry Delugach. Common logic [EB/OL]. [2010-03-07]. <http://cl.tamu.edu>.
- [9] Cycorp. Cyc[EB/OL]. [2010-03-08]. http://www.opencyc.org/faq/opencyc_faq.
- [10] Gellish. Gellish[EB/OL]. [2010-03-08]. <http://sourceforge.net/projects/gellish>.

- [11] Perakath C Benjamin, Christopher P Menzel, Richard J Mayer, et al. IDEF5 method report[M]. Wright-Patterson Air Force Base: Knowledge Based Systems Inc, 1994.
- [12] Michael R Genesereth. Knowledge interchange format [EB/OL]. [2010-03-08]. <http://logic.stanford.edu/kif/dpans.html>.
- [13] W3C. RIF working group[EB/OL]. [2010-03-08]. http://www.w3.org/2005/rules/wiki/RIF_Working_Group.
- [14] W3C. OWL web ontology language[EB/OL]. [2010-03-08]. <http://www.w3.org/TR/owl-features>.
- [15] The University of Manchester. CO-ODE[EB/OL]. [2010-03-08]. <http://www.co-ode.org/downloads/owldoc-server>.
- [16] University School of Medicine. Protégé[EB/OL]. [2010-03-08]. <http://protege.stanford.edu>.
- [17] Enrico Motta. NeOn project[EB/OL]. [2010-03-08]. <http://www.neon-toolkit.org>.
- [18] Li Yuhua, Bandar Z, Mcleod D. An approach for measuring semantic similarity between words using multiple information sources[J]. IEEE Transactions on Knowledge and Data Engineering, 2003, 15(4): 871-882.
- [19] 戴维民. 语义网信息组织技术与方法[M]. 上海: 学林出版社, 2008: 152-159.
- [20] Valtchev P. Construction automatique de taxonomies pour laide la representation deconnaissances par objets [D]. Dissertation; Universite Grenoble, 1999.
- [21] Maedche A, Zacharias V. Clustering ontologybased metadata in the semantic web: proceedings of the 13th ECML and 6th PKDD[C]. Washington: ACM Press, 2002.
- [22] Zhong J, Zhu H, Li Y, et al. Using information content to evaluate semantic similarity in a taxonomy: proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI-95) [C]. Washington: ACM Press, 1995.
- [23] 赵晋巍, 真溱. 本体匹配技术研究概述[J]. 图书情报技术, 2009(11): 6-9.
- [24] 杜小勇, 李曼, 王珊. 本体学习研究综述[J]. 软件学报, 2006, 17(9): 1837-1847.
- [25] 丁晟春, 岑咏华, 顾德访. 基于 Ontology 的语义检索研究[J]. 情报学报, 2006, 24(6): 702-706.
- [26] 叶艳, 杨东, 江志斌. 基于 Ontology 的电子政务流程知识建模与集成管理[J]. 上海交通大学学报, 2006, 40(9): 1549-1555.
- [27] Votisa K, Alexakosa C, Vassiliadis B. An ontologically rincipled service-oriented architecture for managing distributed e-government nodes[J]. Journal of Network and Computer Applications, 2008, 31(2): 131-148.
- [28] Huang Ning, Diao Shihan. Ontology-based enterprise knowledge integration[J]. Robotics and Computer-Integrated Manufacturing, 2008, 24(4): 562-571.
- [29] 梁剑萍, 王术娟, 李吉桂. 基于 Ontology 的数字博物馆通用建设平台研究[J]. 计算机科学, 2008, 34(1): 157-159.
- [30] Paul Buitelaar, Philipp Cimianob, Anette Frank, et al. Ontology-based information extraction and integration from heterogeneous data sources [J]. Int J Human-Computer Studies, 2008, 66(11): 759-788.
- [31] Chang Xiaomeng, Sahin Chang, Terpeny Janis. An ontology-based support for product conceptual design [J]. Robotics and Computer-Integrated Manufacturing, 2008, 24(4): 755-762.
- [32] 周相兵. 用描述逻辑实现语义主题 Web 服务组合的方法[J]. 计算机应用, 2010, 30(10): 2763-2767.
- [33] Zhou Xiangbing, Yang Xiaoping. A model and algorithm of affection semantics web servive component based on ant colony[J]. Journal of Computational Information Systems, 2009, 5(1): 429-437.
- [34] 罗伟, 李陶深. 一种基于本体的个性化搜索引擎模型 [J]. 广西科学院学报, 2006, 22(4): 256-259.
- [35] 周相兵, 余堃, 马洪江, 等. 一种基于云计算的 SaaS 组装方法[J]. 小型微型计算机系统, 31(10): 1942-1953.

(责任编辑: 韦廷宗)