

本体的形式化研究*

冉婕, 漆丽娟

(云南昭通师范高等专科学校 计算机科学系, 云南 昭通 657000)

摘要: 知识本体作为共享概念模型的形式化的规范说明, 在信息技术领域的各个方面有着广泛的应用。本体的表示有多种方法, 本文给出了 OWA(ontology-based Web annotation)、Maedche A 形式化定义、Guarino 和 KAON(Karlsruhe Ontology)4 种典型形式化定义, 并对其进行了对比分析。

关键词: 本体; 形式化; OWL

中图分类号: TP391

文献标识码: A

文章编号: 1674-7720(2012)06-0001-03

The research of formal ontology

Ran Jie, Qi Lijuan

(Department of Computer Science, Zhaotong Teacher's College, Zhaotong 657000, China)

Abstract: Ontology acts as "a formal specification of a shared conceptualization" and is applied to everywhere of information technology domain. There are several ways of ontology's represent. In this paper, we describe OWA (ontology-based Web annotation), Maedche A, Guarino and KAON (Karlsruhe Ontology) four typical formal definitions, and give a comparative analysis.

Key words: ontology; formal; OWL

知识是推理的基础, 是人类文明的成果和发展的基石, 不能想象人类失去了知识将会怎样。知识表示是一种数据结构与控制结构的统一体, 既考虑知识的存储又考虑知识的使用。知识表示是一组描述事物的约定, 把人类知识表示成机器能处理的数据结构。本体是将世界上所有的东西用层次类别组织起来, 它是知识表示的中心内容^[1]。近年来, 本体论被人们引入知识工程领域, 其最终目的是为了解决知识重用和共享。本体不仅是一个可共享的概念模型, 更重要的是: 首先, 本体所描述的概念不能有二义性, 即必须是“明确(explicit)”的; 第二, 本体不仅能被人理解, 还要能被计算机所理解, 所以本体应该被“形式化(formal)”的^[2]。本文介绍了典型的本体的形式化定义, 包括: OWA(Ontology-based Web Annotation)形式化定义、Maedche A 形式化定义、Guarino 形式化定义、KAON(Karlsruhe Ontology)形式化定义等; 并对这四种形式化定义作了简单的对比分析。

1 本体及本体表示

本体是“共享概念模型明确的形式化规范说明”。通过本体(Ontology)可以获取某一领域的知识、本体描述该领域的概念以及这些概念之间的关系^[3]。典型的本体一

般由概念、概念的属性、概念间的关系、概念和属性之间的约束来表示领域知识, 这在本质上决定了本体可以借助成熟的关系数据库技术来完成本体存储、查询等工作。OWL(Web Ontology Language)是 W3C 推荐的语义互联网中本体描述语言的标准^[4]。本体可以采用 OWL 进行形式化编码, 即用 OWL 中定义好的元 ontology 对概念和关系进行形式化描述。

本体的表示有多种: 有仅表示概念的, 有表示概念和属性的框架和语义网络的, 还有能表达丰富语义的基于逻辑的表示。ER 模型和 UML 类图也可认为是本体的表达。当前有逻辑和非逻辑两种主要的本体表达。非逻辑表达直观方便但语义不丰富、不精确, 没有强大的推理能力; 逻辑表达源于对谓词演算, 可以清晰表达所描述的领域的认识。描述逻辑是一阶谓词逻辑的子集, 它只有概念和角色两个基本成分, 较好地与本体的定义相对应。用描述逻辑形式表达的本体语言的两个基本描述是: 原子概念集(其成员用 A 表示)和原子角色集(其成员用 R 表示)。原子概念表示领域中的对象, 原子角色表示对象之间的关系。复杂的描述可以通过以上 2 个概念构造符逐步创建^[5]。本体的形式化描述主要有基于五元组、六元组、七元组和八元组的描述^[6]。参考文献[6]采用函数式描述语言精确定义了本体建模基元, 提出一种具

* 基金项目: 云南省教育厅科学研究基金项目(2011C040)

综述与评论 Review and Comment

有封闭性的本体代数定义;参考文献[7]提出了基于属性文法的本体形式化描述方法;参考文献[8]将本体抽象为外部声明和内部表示两部分,在外部声明中说明本体中的概念与关系,在内部表示中说明概念与关系的语义关系,建立了本体内的简单推理模型。

2 本体的形式化定义

目前本体论的很多定义都具有普遍性,但却很少有准确的数学定义,即很少从形式化角度对本体进行精确的描述。其主要原因是数学定义应该包括所有不同类型的本体论,并且不应与特殊的知识表示方法有关。本体的形式化定义典型的有:OWA形式化定义、Maedche A形式化定义、Guarino形式化定义、KAON(Karlsruhe Ontology)形式化定义等^[9]。

2.1 OWA形式化定义

在基于本体的Web注释OWA的框架中,本体论充当组织Web上语义相关数据的领域知识。与传统的RDB相比,本体由于自身更丰富的语义结构,使之能够描述更复杂的对象。

OWA本体定义是一个六元组: $\{C, A^C, R, A^R, H, X\}$ 。其中, C 表示概念集; A^C 表示每个概念的属性集,概念属性集 $A^C(c_i)$,即概念集 C 中每个概念 c_i 用来表示相同种类的一组对象,并能用相同的属性集进行描述; R 表示关系集,关系 $r_i(c_p, c_q)$ 即关系 R 中的每个关系 r_i 表示概念 c_p 和 c_q 之间的二元关联,并且此关系的实例是一对概念对象 (c_p, c_q) ; A^R 表示每个关系的属性集,关系属性集 $A^R(r_i)$,用于表示关系 r_i 的属性; H 表示概念层次, H 是概念集 C 的概念层次,并是 C 中概念之间的一组父子关系; X 表示公理集, X 中的每个公理是对概念的属性值和关系的属性值的约束,或是对概念对象之间关系的约束。

2.2 Maedche A形式化定义

Maedche A形式化定义是一个五元组: $O = \{C, R, H^c, Rel, A^o\}$ 。其中: C 表示概念集合,其中的一个元素称为一个概念; R 表示关系集合,其中的一个元素称为一个关系; H^c 表示概念的层次或分类层次, $H^c \subseteq C \times C$ 即概念间的分类关系,是一种有向关系, $H^c(C_1, C_2)$ 表示 C_1 是 C_2 的子概念; Rel 表示概念间的非分类关系, $Rel: R \rightarrow C \times C$ 是一个函数, $Rel(R) = (C_1, C_2)$ 也可表示为 $R(C_1, C_2)$; A^o 表示使用某种逻辑语言表达的Ontology公理集^[10]。

2.3 Guarino形式化定义

任何逻辑都隐含自身的本体论,该本体论包含理论假定存在的所有事情,因此逻辑理论是本体中所有实体存在的本体论承诺。Quine从逻辑和哲学的角度研究本体论承诺,规定在逻辑理论强的每个术语都成为该理论的本体。基于Quine的观点,每个逻辑理论都有其自己的显式或隐式的本体,但从知识工程的角度来看,涉及本体的很多知识都能取得轻型本体,把知识库限定在存在于外部本体的术语中,这显然不实际,因此知识工程

中的本体应定义为知识库中的术语和在本体中同一或等同的术语之间的形式化映射。Guarino把本体论承诺表达成在语言和被称为本体的某些事物之间的某种映射。Guarino的本体形式描述如下^[11-12]:

域空间 $\langle D, W \rangle$ 结构,其中 D 是领域, W 是 D 中最大事物状态(或可能世界)的集合。 n 元概念关系 ρ^n ,域空间 $\langle D, W \rangle$ 上的 n 元概念关系是从集合 W 到域 D 中所有 n 元关系集合的映射,即全函数 $\rho^n: 2^D^n$ 。概念化是一个有序三元组 $C = \langle D, W, R \rangle$,其中 R 是域空间 $\langle D, W \rangle$ 上概念关系 ρ^n 的集合。逻辑语言 L 的内涵解释 $\langle D, R \rangle$,其中概念化 $C = \langle D, W, R \rangle$,而函数 $E: V \rightarrow D \cup R$ 是把域 D 的元素赋予语言词汇 V 的常量符号,并把集合 R 的元素赋予词汇 V 的谓词符号。

2.4 KAON形式化定义

德国Karlsruhe大学AIFB学院的知识管理团队在开发Karlsruhe本体框架过程中,经过长期的研讨后,对出现本体论的核心及不同扩展进行正式定义,该定义从本体的角度对知识库作了较全面的形式化定义,对知识库的本体化描述具有普遍代表性,其具体描述如下:

KAON的本体定义是一个五元组: $O := (C, \leq C, R, \sigma, \leq R)$,其中两个不相交的集合 C 和 R 的元素分别被称作概念标识符;偏序 $\leq C$ 称作集合 C 的概念层次;函数 $\sigma: R \rightarrow C^+$ 称作签名;偏序 $\leq R$ 称作集合 R 的关系层次, $r_1 \leq R r_2$ 意味着 $|\sigma(r_1)| = |\sigma(r_2)|$,并且对于每个 $1 \leq i \leq |\sigma(r_1)|$,有 $\pi_i(\sigma(r_1)) \leq C \pi_i(\sigma(r_2))$ 。在实际的应用中,几乎所有的关系都是二元的,故需要定义这些关系的领域和范围。对于关系 $r \in R$,有 $|\sigma(r)| = 2$,则定义此关系的领域是 $dom(r) := \pi_1(\sigma(r))$,范围是 $range(r) := \pi_2(\sigma(r))$ 。

概念的和/或关系之间的关联以及约束能在逻辑语言内进行比较,给出了基于逻辑语言本体定义的公理系统,以便允许使用不同的逻辑语言。假设 L 是逻辑语言,则本体 $O := (C, \leq C, R, \sigma, \leq R)$ 的 L 公理系统是 $A := (AI, \alpha)$,其中集合 AI 中的元素被称作公理标识符; $\alpha: AI \rightarrow L$ 是映射; $A := \alpha(AI)$ 的元素被称作公理。本体 $O := (C, \leq C, R, \sigma, \leq R)$ 的词典是以下结构: $Lex := (S_C, S_R, Ref_C, Ref_R)$,其中两个集合 S_C 和 S_R 的元素分别被称作概念标记与关系标记;如果对于所有的 $c \in C \cap S_C$ 有 $(c, c) \in Ref_C$,则关系 $Ref_C \subseteq S_C C$ 称作概念的词汇参考;如果对于所有的 $r \in R \cap S_R$ 有 $(r, r) \in Ref_R$,则关系 $f_r \subseteq S_R R$ 称作关系的词汇参考。

本体对领域的内涵部分进行形式化,而知识库中包含对概念实例和关系实例的断言,故知识库提供领域的外延部分。知识库是结构 $KB := (C_{KB}, R_{KB}, I, l_C, l_R)$,包含两个集合 C_{KB} 和 R_{KB} ;集合 I 的元素被称作实例标识符;函数 $l_C: C_{KB} \rightarrow \beta(I)$ 称作概念实例化;如果对于所有的 $r \in R$,有 $l_R(r) \in \prod_{c \in \sigma(r)} l_C(c)$,则函数 $l_R: R_{KB} \rightarrow \beta(I^+)$ 称作关系实例化。知识库 $KB := (C_{KB}, R_{KB}, I, l_C, l_R)$ 的实例词典是组对 $IL :=$

综述与评论 Review and Comment

(S_i, R_i) , 其中集合 S_i 的元素被称作实例标记; 关系 $R_i \subseteq S_i$ 称作实例的词汇参考。

本体是对共享的概念化进行形式的规范说明, 是知识工程领域中的一个重要分支。本体的表示有多种, 主要包括逻辑和非逻辑两种主要的本体表达。本文从逻辑表达的角度, 给出了 OWA (Ontology-based Web Annotation)、Maedche A、Guarino 和 KAON (Karlsruhe Ontology) 4 种典型形式化定义, 并对其进行了对比分析。目前, 本体的形式化表示及应用仍是知识工程的一个热点研究领域, 如何用数学的方式更规范地表达本体, 将是下一步研究的重点。

参考文献

- [1] 吴强, 刘宗田, 强宇. 基于本体的知识库推理研究[J]. 计算机应用研究, 2005, 21(13): 50-52.
- [2] 张其文, 臧凤奎, 李明. 本体语言的联系及其逻辑基础分析[J]. 科学技术与工程, 2009, 9(4): 950-953.
- [3] 王乐, 张建军. OWL 本体存储的分析与应用[J]. 科学技术与工程, 2008, 8(7): 1831-1834.
- [4] OWL Web ontology language guide recommendation[EB/OL] (2004-02-10). [2011-12-10] <http://www.w3.org/TR/2004/RRC-owlguide-20040210>.
- [5] 王真星, 吕滕. 基于描述逻辑的本体导航[J]. 计算机工程, 2005, 31(11): 28-29.
- [6] 王俊华, 左万利, 赫枫龄, 等. 本体定义及本体代数[J]. 吉林大学学报(理学版), 2010, 48(6): 1001-1007.
- [7] 宫慧颖, 汪美玲, 刘磊. 基于属性文法的本体形式化描述方法及其应用[J]. 吉林大学学报(信息科学版), 2005, 23(2): 199-204.
- [8] 张大志, 刘磊. 一种本体的形式化描述方法及其应用[J]. 吉林大学学报(信息科学版), 2004, 22(1): 74-78.
- [9] 程显毅, 刘一松. 面向智能体的知识工程[M]. 北京: 科学出版社, 2008.
- [10] MAEDCHE A. Ontology learning for the semantic Web [M]. [S.l.]: Kluwer Academic Publishers, 2002.
- [11] GUARINO N. Semantic matching: formal ontological distinctions for information organization extraction and integration[C]. In: Pazienza M T, eds. Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology, Springer Verlag, 1997: 139-170.
- [12] GUARINO N, WELTY C. A formal ontology of properties [C]. In: Dieg R, Corby O, eds. the Proceedings of the 12th International Conference on Knowledge Engineering and Knowledge Management (EKAW'2000), Springer Verlag, 2000: 97-112.

(收稿日期: 2011-12-29)

作者简介:

冉婕, 女, 1975 年生, 副教授, 硕士研究生, 主要研究方向: 本体构建及语义检索。

漆丽娟, 女, 1977 年生, 讲师, 硕士研究生, 主要研究方向: 人工智能。