

# 本体与语义网

史程

(大连海事大学 信息科学技术学院 大连 中国 116026)

**摘 要** 本文概述了本体的概念、本体的描述语言和本体的构建方法,介绍了本体的用途,并简要讲述了语义网的体系结构和本体在语义网中的应用。

**关键词** 本体; 语义网

中图法分类号 TP18 DOI号: 10.3724/SP.J.1016.2014.01229

## Ontology and Semantic Web

SHI Cheng

(Information Science and Technology College, Dalian Maritime University, Dalian 116026, China)

**Abstract** In this paper, the concept, description language, building methods of the ontology are briefly described, and introduces the application of the ontology. Besides, we briefly introduce the architecture of the Semantic web and the application of ontology in Semantic web.

**Key words** Ontology; Semantic Web

## 1 引言

本体论 (Ontology) 是一门重要的新兴学科,现已成为知识工程、自然语言处理、信息系统、智能系统集成和知识管理等多个领域的热门研究方向。它为人们及广泛异构的应用系统提供共同的领域知识理解,并为第三代互联网 (语义网, Semantic Web) 中基于内容的知识获取、互用和交流提供高质量的保证。

本文将试从本体的概念、描述语言、构建方法及其用途等方面做一系列的概要介绍,并简述语义网的结构体系和本体在语义网中的基础作用。

## 2 本体

### 2.1 本体的概念

本体是对客观的事物以一种形式化的、客观的并且系统化的方式进行描述。本体由哲学领域发起,对现实世界的客观事物进行本质化的描述。它在哲学中的定义为“对世界上客观存在物的系统地描述,即存在论”,是客观存在的一个系统的解释或说明,关心的是客观现实的抽象本质<sup>[1]</sup>,现在较多的翻译为本体论。后来随着在人工智能、计算机以及网络领域中的应用发展,其定义也被融入了许多新的内容。其中最著名、被引用最为广泛的定义是由 Gruber 提出的:“本体是概念化的明确的规范说明”。Studer 对本体诸多定义进行概括分析后认为,本体论的概念包括四个主要方面<sup>[2]</sup>:

(1) 概念化 (conceptualization): 客观世界现象的抽象模型,其表示的含义独立于具体的环境状

态;

(2) 明确 (explicit): 概念及它们之间联系都被精确定义;

(3) 形式化 (formal): 精确的数学描述, 计算机可读;

(4) 共享 (share): 本体中反映的知识是其使用者共同认可的, 是相关领域中公认的概念集, 它所针对的是团体而不是个体。

本体的目标是捕获相关领域的知识, 提供对该领域知识的共同理解, 确定该领域内共同认可的词汇, 并从不同层次的形式化模式上给出这些词汇(术语)和词汇相互关系的明确定义。

## 2.2 本体的描述语言

本体需要某种语言来对概念化进行描述, 按照表示和描述的形式化的程度不同, 可以将本体分为完全非形式化本体、半非形式化本体、半形式化本体和严格形式化的本体。本体可以用自然语言来描述, 也可以用框架、语义网络或逻辑语言等来描述。但要达到清晰的、形式化的概念描述, 本体语言就应满足以下要求: 良好定义的语法, 良好定义的语义, 有效的推理支持, 充分的表达能力, 表达的方便性。<sup>[3]</sup> 本体描述语言随着本体研究的升温, 不断推陈出新, 主要有 RDF 和 RDFS、OIL、DAML、DAML+OIL、OWL、KIF、SHOE、XOL、OCML、Ontolingua、CycL、Loom 等。

### 2.2.1 XML

XML (Extensible Markup Language) 是针对包含结构化、半结构化信息的文档而设计的一种标记语言。XML 已经成为了 Web 上数据表示和交换的事实标准, 是应用之间或者机器之间共享数据的一种有效方式。XML 的可扩展性是 XML 区别其他标记语言的最基本特征, 其核心在于以一种标准化的方式来建立数据表示的结构, 而将具体标记的定义留给了用户。但是, XML 并不能解释它标记的含义, 并且 XML 模式只能对 XML 的语法合法性进行验证, 而不能区分 XML 属性和元素在含义上的不同。因此对于同样的信息内容, 我们可以将其映射为多种不同的 XML 结构, 同时对同一种 XML 结构, 也可以存在多种不同的解释, 而相同的应用也可能对不同的 XML 做出同样的解释。

### 2.2.2 RDF 与 RDFS

RDF 定义了一种用以描述资源及其相互关系的简单模型, 是语义互联网实现的关键技术之一, 也是语义信息描述的有效手段。其基本数据模型包

含三类对象: 资源、属性和陈述, 资源之间的关系通过属性和值来描述, 描述特定资源特定属性的值, 就构成 RDF 中的一个陈述, 通常可以用三元组  $\langle \text{subject}, \text{predicate}, \text{object} \rangle$  描述。其中, 被描述的资源称为 subject, 描述资源的属性称 predicate, object 则是属性对应的值。RDF 建立在 XML 和 URI 的基础上。

RDF 通过属性和值描述了资源以及资源之间的关系, 但并没有提供描述这些属性及属性间关系的机制。RDF Schema 提供了这种表达机制, 它描述了 RDF 中 properties 的使用规则, 为 RDF 定义了领域字典, 并用类型层次结构来组织该字典, 从而构成完备的语义空间。RDF 的数据模型实质上是一种二元关系的表达, 由于任何复杂的关系都可以分解为多个简单的二元关系, 因此 RDF 的数据模型可以作为其他任何复杂关系模型的基础模型。

RDF Schema 在提供了简单的机器可理解语义模型的同时, 为领域化的本体语言提供了建模基础, 并使得基于 RDF 的应用可以方便地与这些本体语言所生成的本体进行合并。RDF 的这一特性使得基于 RDF 的语义描述结果具备了可以和更多的领域知识进行交互的能力, 也使基于 XML 和 RDF 的 Web 数据描述具备了良好的生命力。

### 2.2.3 OWL

OWL 目前是本体的标准描述语言。它是结合了 DAML+OIL 应用经验而改进的修订版, 建立在 RDF 基础上, 以 XML 为书写工具。主要用来表达需要计算机应用程序来处理的文件中的知识信息, 而不是呈递给人的知识。OWL 能清晰的表达词表中各词条的含义及其之间的关系, 这种表达被称为本体。OWL 相对 XML、RDF 和 RDF Schema 拥有更多的机制来表达语义, 从而 OWL 超越了 XML、RDF 和 RDF Schema 仅仅能够表达网上机器可读的文档内容的能力。<sup>[4]</sup>

## 2.3 本体的构建方法

本体的构建主要有手工构建、复用已有本体和自动构建本体等多种方式, 由于构建本体的领域范围、设计标准与原则等不相同的特点, 本体的构建工作没有统一的实现标准。构建本体的目的有多种多样, 构建本体的步骤和过程也各不相同。一般的来说, 本体的构建应该遵循明确性和客观性、完全性、一致性等原则<sup>[1]</sup>。一般说来, 建构一个知识领域的本体, 包括以下 6 个步骤:

(1) 确定本体的领域和范围

首先,要明确建构的本体将覆盖的专业领域、应用本体的目的、作用及其系统开发、维护和应用的对象。

#### (2)列举知识领域中重要的术语、概念

在创建本体的初始阶段,尽量列举出系统想要陈述的或准备向用户解释的所有概念,不必考虑概念之间语义的重叠及表达方式(类、属性、实例)。

#### (3)建立本体框架

上一步生成的知识领域中的大量概念是一个没有组织结构的词汇表,需要按一定的逻辑规则,将其分组,构成不同的工作领域,并对同一工作领域内的概念相关性和重要性进行评估,选出关键技术术语,尽可能准确而精炼地表达出该领域的相关知识,形成该领域知识的框架系统。

#### (4)设计元本体,重用已有的本体,定义领域中概念之间的关系

元本体是元概念的本体,其术语用于定义本体中的高层次的抽象概念,如实体、关系、角色等。设计元本体时,一要尽量作到领域无关性;二要包含的元概念尽可能少。一个概念可采用元本体中定义的元概念进行定义,或采用本体中已被定义的概念进行定义或重用已有的本体。

除了定义概念之外,还要定义概念之间的关系。这些关系不仅涉及同一工作领域中的概念,而且还与其它工作领域的概念相关联。

#### (5)对领域本体进行编码、形式化

选用合适的本体描述语言,对上述建立的本体进行编码,形式化。现有本体描述语言约 28 种,大都基于一阶逻辑,也有基于描述逻辑。常用的本体语言有: Ontolingua、CycL、Loom 等。本体模型实现形式化可提供比自然语言更严格的格式,能增强机器的可读性,便于交换及本体模型自动逻辑推理及检验。

#### (6)对本体进行检验和评价

本体在形式化以后,是否满足用户需求,是否符合本体的建构准则,是否术语、概念定义清晰,是否关系定义完整等,都要在本体建构后进行检验和评估。

### 3 本体的用途

构建一个本体可以解决以下问题:

(1) 在用户间或软件代理间达成对于信息组织结构的共同理解和认识<sup>[5]</sup>。假如有若干包含医药信

息或提供医药电子商务服务的 Web 站点,这些 Web 站点共享相同的底层本体,那么计算机代理就可以准确、没有歧义的抽提和集成这些来自不同站点的信息,代理软件可以利用这些集成的信息来回答用户的检索式或向用户提供数据。

(2) 可以复用专业领域知识。譬如,中医各科对诊断技术的多方面理解。

(3) 使专业领域内的假设变得更加明确。对于必须理解该领域术语的新用户很有帮助。

(4) 从操作性知识中分离出领域知识。

(5) 分析专业领域的知识。在进行复用现有本体和扩展这些本体的尝试中,对术语规范地分析是极有价值的<sup>[3]</sup>。

因为本体是以机器可以理解的形式化语言来描述知识,目的是从根本上解决人与机器、机器与机器之间的信息、知识交流障碍,因此,它几乎可以应用在存在有信息、知识交流的各种领域,如:医学、电子商务、房地产、金融财务、制造业和工商业的供应链、电子教学、图书馆等,最大的用途当属各种信息系统和语义网的构建。

### 4 本体与语义网

语义网的核心内容是建立一个明确的语义空间。其中,需要解决的关键问题就是语义的表达和结构化的信息集合及推理规则。XML(Extensible Markup Language)可扩展标记语言和 RDF(Resource Description Framework)资源描述框架技术解决了这一问题。XML 是一种用于定义标记语言的工具,其内容包括 XML 声明、定义语言语法的 DTD(Document Type Declaration 文档类型定义)、描述标记的详细说明以及文档本身,它提供了灵活、通用、丰富的结构化信息表示方式,是整个语义网的基石。RDF 提供语义信息和推理规则的表达方式,是语义网表达语义的关键。理论上构建的语义网有 7 个层次组成,如图 1 所示。

Unicode 和 URI 层:保证用户使用国际化、通用化的字符集,同时也可以实现多国语言的混合存储和使用。

XML+NS+XML Schema 层:XML 是一种允许自定义标记的、描述 Web 文档和数据的结构化描述语言。命名空间 NS 为 XML 文档中的结构化标记提供上下文环境。XML Schema 在标记的使用上和文档结构上,为 XML 文档提供了明确的语义限制。

**RDF/RDF Schema 层:** RDF 定义一种用以描述资源及其相互关系的简单模型,是语义信息描述的有效手段。其基本数据模型包含 3 类对象:资源、属性和陈述;资源之间的关系通过属性和值来描述。RDF Schema 提供属性及属性间关系的表达机制,描述 RDF 的使用规则,定义领域字典,并用类型层次结构来组织该字典,构成完备的语义空间。

**Ontology Vocabulary 层:** 在交流/通讯中扮演语义沟通的角色,用于描述的、概念化的显示说明。

**Logic 层、Proof 层和 Trust 层:** 是语义表达的高级要求。其中,Logic 层提供推理规则的描述手段,Proof 层通过运用这些规则进行逻辑推理和求证,Trust 层则负责为应用程序提供一种机制,且对“是否信任”给出的论证,作出结论;**Digital Signature** 位于层次模型的右侧,并且贯穿于中间的 4 层。**Digital Signature** 是一种基于互联网的安全认证机制,通过它可以鉴别信息的来源和信息的安全性,用户据此决定是否接受该信息。

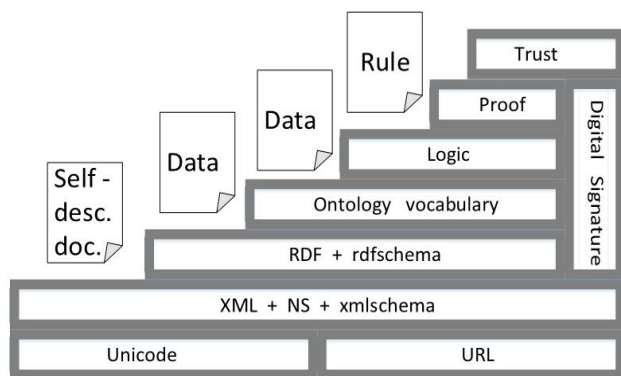


图 1 语义网结构图

本体具有良好的概念层次结构和对逻辑推理的支持,在语义网中将会有广泛的应用,主要包括如下几个方面:

(1) 电子商务网站: 本体有助于商家和客户之间实现基于机器的通信,使得市场的垂直集成成为可能,并使得商品的描述能在不同的市场复用? 例如: Yahoo shopping 和 Amazon1 com 都提供了基于本体的更为强大的信息搜索和浏览功能;

(2) 搜索引擎: 本体能帮助实现跨越目前基于关键词的搜索,并可找到包含语法不同但语义相似的词或句的页面;

(3) Web 服务: 本体可提供易于被智能代理所理解的语义丰富的服务描述。例如美国的 SHOE (Simple HTML Ontology Extensions, 利用本体论将 HTML 扩展为语义网代码) 工程和欧洲的 IST 工程(本体被用来促进企业网之间的知识交流) 等。

## 5 结束语

本文介绍了本体的相关知识,探讨了本体语言建模的方法和语言,讲述了语义网的体系结构并分析了本体在语义网中的基础作用。

## 参考文献

- 1 邓志鸿,等. Ontology 研究综述[J]. 北京大学学报(自然科学版), 2002; 38(5): 730-738
- 2 Studer R, Benjamins V R, Fensel D. Knowledge Engineering, Principles and Methods[J]. Data and Knowledge Engineering, 1998; 25(12): 161-197
- 3 Michael D. Ontology Building: A Survey of Editing Tools1/ OL2. <http://www.xml.com/pub/a/2002/11/06/ontologies.html>
- 4 <http://iswc2004.semanticweb.org/>
- 5 周竞涛,王明微. XML+ RDF——实现Web数据基于语义的描述[J/OL]. <http://www-900.ibm.com/developerWorks/cn/xml/x-xmlrdf/index.shtml>
- 6 刘炜. 元数据和知识本体[J]. 图书馆杂志, 2004, 23 (6): 50-54.
- 7 李健康,张春辉. 本体研究及其应用进展. 图书馆论坛, 2004; 24(6)
- 8 刘植惠. 本体与语义网. 重庆图情研究, 2006; 3