

《智能信息处理》课程考试

## 本体构建方法的研究

吴继萌

考核	到课[10]	作业[20]	考试[70]	课程成绩[100]
得分				

2020 年 12 月 7 日

# 本体构建方法的研究

吴继萌

**摘要：**本体构建的方法有很多，例如：TOVE 法，METHONTOLOGY 法，骨架法，KACTUS 法，七步法和基于叙词表的领域本体构建。构建本体是一项复杂的系统工程，需要科学规划，精心实施。本文从本体的分类、常用的本体构建方法、本体构建工具和本体构建方法存在的问题这四个方面对本体的构建进行了阐述。

**关键词：**本体、本体构建的方法

## Research on Ontology Construction Method

Wu Jimeng

**Abstract:** There are many methods of ontology construction, such as: TOVE method, METHONTOLOGY method, skeleton method, KACTUS method, seven-step method and thesaurus-based domain ontology construction. Constructing ontology is a complex system engineering that requires scientific planning and careful implementation. This paper expounds the construction of ontology from four aspects: classification of ontology, common ontology construction methods, ontology construction tools and problems in ontology construction methods.

**Keywords:** ontology, method of ontology construction

## 1 引言

随着语义 Web 和信息大爆炸的到来，大规模抽取并表示信息的系统研究变得越发重要。近年来，本体学习逐渐为研究人员熟知，原因是获取信息较为简单且能提供可共享的高级结构。此外，由于本体能够概念化地描述事物的特征并在它们之间建立逻辑关系，这种结构化的可共享信息被广泛应用，目前主要集中在信息检索、人工智能、信息抽取、异构信息系统集成、语义 Web 等领域。但是，作为一种较为抽象的概念表达方式，本体在具体应用中受到一些挑战：本体在描述庞大的信息并对其概念化时难度较大；随着本体应用领域实体的多样化，本体描述语言相应也变得更需要具有兼容性。

20 世纪 90 年代，人工智能领域的研究者逐步认识到，知识获取是开发人工智能系统的关键，并且开

始采用类似本体的计算模型实现自动推理。80 年代，人工智能领域正式将“Ontology”作为术语，表示建立客观界抽象模型的一种理论或者知识系统的组成部分。90 年代，本体研究呈现井喷式增长；进入 21 世纪，作为语义网的关键支撑技术，本体广泛应用于自然语言处理、信息检索、知识系统等诸多领域。在这一发展过程中，为了分享本体构建经验，推动本体开发实践，提出了许多本体构建方法。经典本体构建方法都有自己的特点和适用领域，再加上本体构建本身也没有统一标准，因此难以在不同领域本体的构建中保持一致。

## 2 研究现状

国外：国外对本体建模作了大量研究并将其运用于知识工程领域。主要代表为：① 万维网联盟 W3C 的研究；② 德国卡尔斯鲁厄大学基于

本体的知识门户和语义门户的研究；③ 美国斯坦福大学的知识系统实验室对本体建模工具和本体应用层面的研究。

国内：国内进行本体研究的主要有三支科研力量。1) 中国科学院计算所、数学所、自动化所的若干实验室，代表人物是陆汝铃院士等人。2) 哈尔滨工业大学计算机系，代表人物是王念滨博士。3) 浙江大学人工智能研究所，代表人物是博士生导师高济教授。

### 3 本体

#### 3.1 本体的定义

本体（Ontology）的概念源自于哲学领域，在哲学中的定义为“对世界上客观事物的系统描述，即存在论”。哲学中的本体关心的是客观现实的抽象本质。而在计算机领域，本体可以在语义层次上描述知识，可以看成描述某个学科领域知识的一个通用概念模型。德国学者Studer在1998年给出了本体的相关定义“本体是共享概念模型的形式化规范说明”。这个定义包含了四层含义：即共享（share）、概念化（Conceptualization）、明确性（Explicit）和形式化（Formal）。

（1）共享：指本体中体现的知识是共同认可的，反映在领域中公认的术语集合。

（2）概念化：指本体对于事物的描述表示成一组概念。

（3）明确性：指本体中全部的术语、属性及公理都有明确的定义。

（4）形式化：指本体能够被计算机所处理，是计算机可读的。

本体通常用来描述领域知识。我们可以这样理解它：本体是从客观世界中抽象出来的一个概念模型，这个模型包含了某个学科领域内的基本术语和术语之间的关系（或者称为概念以及概念之间的关

系）。本体不等同于个体，它是团体的共识，是相应领域内公认的概念集合。



图 1 领域本体示例

#### 3.2 本体的描述语言

本体描述语言实质上是一种可以对本体进行描述的代码，有助于更好的描述领域内的本体。本体语言要满足以下要求：对语法及语义要有很好的定义、要支持逻辑推理、要具备一定的表达能力、要方便用户使用。目前本体的语言很多，主要分为 4 大类：非形式化语言、半非形式化语言、半形式化语言、形式化语言。早期的本体语言有 RDF（S）、SHOE、OIL、DAML-ONT、DAML+OIL 等。下面主要介绍 RDF（S）。RDF 即资源描述框架<sup>[2]</sup>，它通常被成为一种“语言”，实际上是一个数据模型（Data-Model）。RDF 对语义的描述比较充分，可借助简单的二元关系描述相关资源信息。RDF 的基本概念是：资源、属性、书目本体构建及语义检索和陈述。资源可以看作是一个对象，任何被描述的事物都是一个资源，如作者、书籍、出版社等。任意一个资源都能够用一个唯一的 UIR 来标识；属性是一类描述资源之间的关系的特殊资源，也用 UIR 标识，从而避免了分布式数据的命名冲突；

陈述是对资源的所有属性进行描述，一个陈述就是描述一个特定资源特定属性的值，是一个形如“对象—属性—值”的三元组（Triple）。每个三元组由三个独立的部分构成：主语、谓语和宾语，分别与对象、属性、值相对应。在 RDF 中任何资源之间的关系都可以用这样的三元组来描述。

### 3.3 本体的分类

通常研究领域不同，本体的分类方法也不尽相同，在某特定领域中构建本体的目的能确定本体的类型，进而确定本体具备的功能。依据本体表达的概念模型所在的层次，可对本体做如下分类：知识本体：将知识表示抽象化，将领域概念以及它们之间的关系规范化，重点研究内容是实现语言对知识的表达描述。通用本体：从认识论出发，表达抽象的概念。如初始概念的时间、空间、事物、数量、状态、属性，以及这些初始状态组成的下位概念。这类本体主要用在应用常识知识。领域本体：抽象出特定领域内的知识，对其中的概念和它们之间的关系、在领域内进行的活动、这些活动相关领域的领域知识和原理研究 等进行描述。应用本体：描述与领域内特定任务相关的概念和它们之间的关系。有关任务的描述要依据特定的领域知识本体。以上介绍的四类本体目前已经得到了国际的广泛认可。

关于本体的研究非常广泛，最为常用的分类方法是根据本体应用主题，将这些为数众多的本体划分为五种类型：领域本体、通用或常识本体、知识本体、语言学本体和任务本体。而依据本体的层次和领域依赖度，Guarino 等人将其分为四类：顶层本体、领域本体、任务本体和应用本体。

（1）顶层本体：研究通用的概念以及概念之间的关系，如空间、时间、事件、行为等，与具体的应用无关，完全独立于限定的领域，因此可以在较大范围内进行共享。

（2）领域本体：研究的是特定领域内概念及概念之间的关系。

（3）任务本体：定义一些通用任务或者相关的推理活动，用来表达具体任务内的概念及概念之间关系。

（4）应用本体：用来描述一些特定的应用，既可以引用领域本体中特定的概念，又可以引用任务本体中出现的概念。

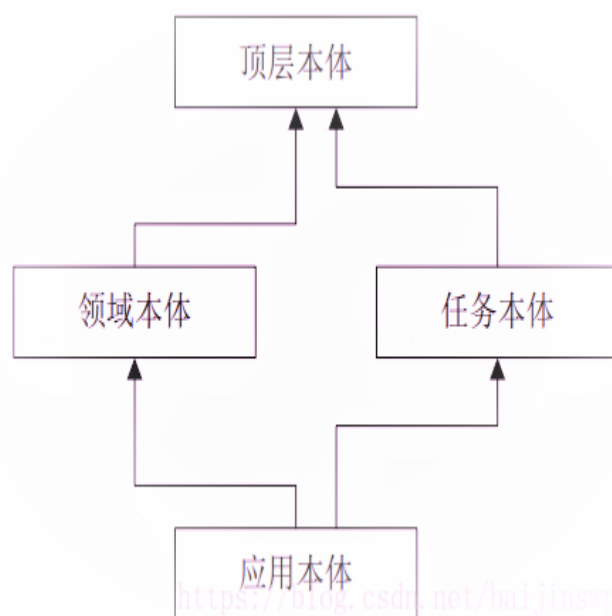


图2 本体

## 4 本体的构建和存在的问题

### 4.1 本体构建的方法

本体的构建方法多种多样，常见的有如下几种。

1、TOVE 法：TOVE 是指多伦多虚拟企业（Toronto Virtual Enterprise）专门用于构建 TOVE 本体（关于企业建模过程的本体）。

2、METHONTOLOGY 法：专用于构建化学本体（有关化学元素周期表的本体）。

3、骨架法：专门用于构建企业本体，建立在企业本体基础之上，是相关商业企业间术语和定义

的集合。

4、KACTUS 工程法：是基于 KACTUS 项目而产生的，KACTUS 是指关于多用途复杂技术系统的知识建模工程。

5、七步法：斯坦福大学医学院开发的七步法，用于领域本体构建。七步骤为：确定领域本体的范畴；复用现有的本体；列出领域内的术语；定义类和类的等级关系；定义类的属性；定义属性的分面；填充实例。

6、基于叙词表的领域本体构建：叙词表又称为主题词表，是一种语义词典，由术语及术语之间的关系组成，能够反映某学科领域的语义相关概念。由于叙词表包含丰富的领域概念和一定的语义关系，在表达和知识结构上与本体有着天然联系，包含了本学科比较完整的术语，因此，国内外很多学者都在尝试将叙词表转换为本体。国内目前基于叙词表已经转化成型的本体原型有《国防科学技术叙词表》和《中国农业科学叙词表》。

## 4.2 本体构建的工具

本体开发是一项庞大的工程,需要借助开发工具来完成本体的构建任务。本体构建工具主要用于本体的开发,多数工具都具有编辑、图示、自动将系统内容转换为数据库、自动转换置标语言等功能。目前常用的本体构建工具主要分为两类:可视化手工构建工具和半自动化构建工具。

### 1、可视化手工构建工具

可视化手工构建工具主要有 Protégé、Apollo、Web Onto、Web ODE 和 Onto Edit 等,这类工具通常为用户提供可视化界面,用户可以通过简单的操作完成本体的构建。

a、Protégé。Protégé 具有图形化的用户界面,操作简单便捷,提供详细的帮助文档,支持模块化设计[24]。并且支持 DAML+OIL 和 OWL 语言,可利用 RDF、RDFS 和 OWL 等本体描述语言在系统外对本体

进行编辑和修改。Protégé 由于其开放源码、支持中文编辑而深受国内学者青睐。但 Protégé 最大的缺陷在于不能批量导入数据,构建大规模本体费时费力,手工输入错误率比较高,效率较低。

b、Apollo。Apollo 采用 Java 语言,支持知识模型中的所有原语:本体、实例、类、功能和关系,可在编辑过程中完成一致性检测。但其在本体语言的支持能力、逻辑推理能力和可扩展性等方面还存在问题。

c、Web Onto。Web Onto 是基于 OCM L 推理引擎的知识模型,提供多重继承、锁机制,支持用户浏览、构建和编辑本体,提供定制数据表示类型的选项和客户端 API,便于检索信息。但是 Web Onto 不提供源代码,扩展性较差。

d、Web ODE。Web ODE 是本体设计环境(ODE)的升级版,可以通过 Java、RMI、COBRA、XML 等技术实现,支持 METHONTOLOGY 本体构建方法,可通过定义实例集来提高概念模型的重用性,具有很大的灵活性和扩展性。

e、Onto Edit。Onto Edit 是一个支持用图形化的方法实现本体开发和管理的工程环境,提供对本体的并发操作,支持 RDFS、DAML+OIL 语言,并且支持多种插件,具有良好的扩展功能。然而这几种构建工具都没有自动或者半自动获取知识和维护能力,所构建的本体兼容性差,在异构系统中难以复用。

## 4.3 本体构建过程中存在的问题

### 1、手工构建

TOVE 法、骨架法、IDEF-5 法、METHONTOLOGY 法、SENSUS 法、KACTUS 工程法、七步法。本体的构建主观性太强,且比较随意,缺少科学管理和评价机制。主要问题:

- 1) 需求描述不充分和建设过程的无计划性
- 2) 建设过程缺少规范性

- 3) 成果没有评价标准
- 4) 忽视本体的共享和重用

## 2、复用已有的本体

基于叙词表和基于顶层本体的构建方法均属于复用已有本体的半自动构建方法。复用已有的本体,可以获得领域知识以及概念关系,使得本体构建有一个很好的起点。

目前可复用的本体资源主要有:① 叙词表资源,如中国农业科学叙词表、国防科学技术叙词表等;② 顶级本体,如 Cyc、SUMO、WordNet、FrameNet 等;③ 数据库资源;④ 在线本体库,如 Ontology Engineering Group 和 DAML。

瓶颈问题:

- 1) 很少有现存的不经修改就能被复用的本体
- 2) 有不少领域没有可供利用的本体资源。
- 3) 本体复用带来了不同本体匹配的问题
- 4) 有些本体资源改造起来需要大量的投入,改造否值得待研究。

## 3、自动构建本体

自动构建本体是目前的一个研究热点。研究者借鉴知识获取的相关技术,有基于自然语言规则的方法和基于统计分析的机器学习方法。

- 1) 利用机器学习会产生大量的噪音数据
  - 2) 缺乏必要的语义逻辑基础
  - 3) 抽取的概念关系松散且可信度无法得到很好的保障。
- 4、利用自然语言处理,概念间潜在关系的分析则需要依赖复杂的语言处理模型。

尽管机器学习应用于本体自动构建有巨大的潜力,但是距离良好的可理解性尚有很大的距离,随着研究的深入这种状况应该有望得到改善。

## 5 小结

本体是客观世界的抽象模型,它以计算机能够理解和计算的方式表示。本文揭示了本体的本质,还指

明了构建本体的思路 and 方向。领域包含大量的概念和对象,要迅速建立一个全面的领域本体是不现实的,本文只是简单的介绍了一下本体的构建方法,之后还有长期的本体完善与操作阶段,其中有大量工作有待研究和实践。

## 参考文献

- [1]本体 (Ontology) 概述
- [2]【领域本体】领域本体的构建方法概述
- [3]王向前,张宝隆,李慧宗.本体研究综述[J].情报杂志,2016,35(06):163-170.
- [4]段宇锋,黄思思.本体构建方法研究[J].情报杂志,2015,34(11):139-144.
- [5]张文秀,朱庆华.领域本体的构建方法研究[J].图书与情报,2011(01):16-19+40.
- [6]马旭明,王海荣.本体构建方法与应用[J].信息与电脑(理论版),2018(05):33-35+38.