

# 服务于本体逻辑理论的本体定义

吕 鲲, 曾庆良, 万丽荣, 曾庆田, 刘志海

(山东科技大学机械电子工程学院, 青岛 266510)

**摘要:** 本体概念过分模糊和庞大, 不利于构建、共享和重用。该文从当前的研究和应用需求出发, 给出一个强调论域的本体定义, 是完善本体逻辑理论的基础。总结已有的研究成果, 基于统一本体层次化和模块化的思想, 提出子本体、原子本体以及本体的递归定义, 为本体建立科学的分类体系和一致的逻辑理论打下了基础。

**关键词:** 本体工程; 子本体; 原子本体; 递归定义

## Definition of Ontology Servicing Ontological Logical Theory

LV Kun, ZENG Qing-liang, WAN Li-rong, ZENG Qing-tian, LIU Zhi-hai

(College of Mechanical and Electric Engineering, Shandong University of Science and Technology, Qingdao 266510)

**【Abstract】** In the developing process of ontology engineering from pure theory to application research, the definition of ontology is too obscure and enormous to facilitate to construct, share, and reuse. Driven by the requirements on current research and application, a new definition of ontology that focuses on a domain is proposed. To unify the ideas of ontological hierarchy and modularization. Sub-ontology, atom ontology, and the recursive definition of ontology are made under summarization of the existing results. These efforts build a foundation for ontology to form the scientific taxonomic system and consistent logical theory.

**【Key words】** ontology engineering; sub-ontology; atom ontology; recursive definition

所谓本体工程是指通过概念分析、建模, 把现实世界中的实体抽象为一组概念与概念之间的关系的理论和方法。它已经成为计算机科学和知识工程的一个重要分支, 在知识密集型领域(如计算机集成制造、基于网络的协同产品开发)的知识管理和信息交互方面显示出巨大的潜力。开发本体的最初目的是为了共享和重用知识, 但随着本体应用的广泛和深入, 人们对本体提出了更高的要求, 如出于安全和保密的考虑, 要求本体局部隐藏和重用<sup>[1]</sup>。本体的模块化和层次化是满足这一需求的有效方法, 这需要一个良构的本体概念作支撑。本文从应用需求出发, 基于本体的模块化和层次化思想, 通过科学的分析, 给出了本体的几个重要定义。

### 1 本体的基本定义

本体概念是本体工程的基石, 本体工作者对本体概念的思索一刻也没有停止过<sup>[2]</sup>。但是许多本体定义过分抽象、庞大而且混乱, 甚至有些是相互矛盾的<sup>[3]</sup>, 这不仅不利于本体的理解和构建, 也不利于本体的重用和维护, 已经阻碍了人们对本体工程的学习和研究。经过前期的研究, 人们提出了本体的一个新定义。这里复述如下:

**定义1** 本体是关于一个论域的、明确的、形式化的、规范的、可共享的概念模型。

这一概念与传统概念的不同在于, 把本体看成是一种特殊的对象, 是关于一个论域的概念模型, 是一个具体的模型而不是一套抽象的理论。它具有明确性、形式化、规范性、可共享等基本特征。依据这一定义, 本体的规模取决于论域的大小, 人们可以根据需要构造合适的本体; 应用需求是变化的, 创建本体应该考虑可维护性和可扩展性; 论域包含问题域、主题域、应用域、兴趣域以及领域等概念, 体现了本体的主观性和客观性的统一。

### 2 本体的模块化和层次化思想

一个大型本体不仅难以共享和重用, 也难以构造和维护。为了实现本体的可重用性和可用性的折中, 模块化是比较理想的方法, 它体现一个领域的知识模型是由多个本体组成的这一事实。

#### 2.1 关于本体的包

Bao等人<sup>[1]</sup>为了改善本体的可重用性和选择性隐藏功能, 提出了本体的模块化思想。他们把本体的一个相对联系紧密的部分打包为一个模块, 称为本体的一个包。并对传统的描述逻辑进行了扩展, 开发了基于包的描述逻辑语言, 给出了“包”的语义: 每个典型的包都由一组密切相关的术语以及这些术语间的关系组成; 一些包可以嵌套在另一些包中, 构成包的层次化结构; 包的可见性由范围限制符来控制, 如公有和私有。图1是生物本体的组织结构和语义结构。

组织结构中的每一个虚框对应一个“包”, “包”显然是本体的一个可以独立存在的组件, 并满足本体的条件和特征, 因此, 是事实上的本体。对应图1可以看出, “宠物”本体是“动物”本体的子本体, “动物”本体是“生物”本体的子本体。注意, “我的宠物”并非“宠物”的子本体, 因为它包含

**基金项目:** 国家自然科学基金资助项目“基于领域本体和过程挖掘的用户知识需求获取的关键技术研究”(60603090); 山东省教育厅基金资助项目“面向协同产品开发的跨领域本体建模与集成技术研究”(J06P15)

**作者简介:** 吕 鲲(1966 -), 男, 副教授、在职博士研究生, 主研方向: 先进制造技术, 知识工程; 曾庆良, 博士、博士生导师; 万丽荣, 教授、博士; 曾庆田, 副教授、博士; 刘志海, 讲师

**收稿日期:** 2007-12-26 **E-mail:** lvkun228@126.com

不是动物的宠物“机器人”。

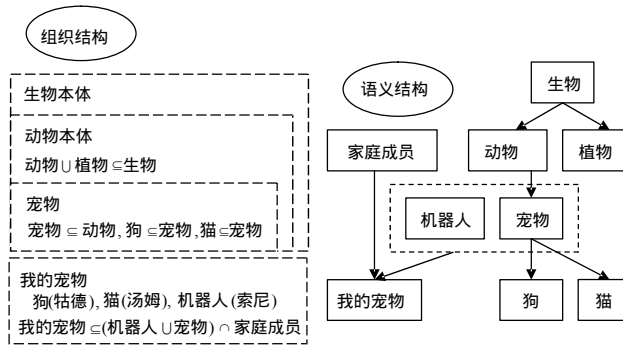


图1 组织结构和语义结构

## 2.2 关于本体的分面

文献[4]提出了“分面”理论(采用“分面”翻译该词是为了与分类法中的分面相对应,可译为“方面”,这类类似于McCarthy和Guha等提出的“语境”思想),目的也是为了共享和重用知识的方便。他们认为“分面”是对概念模型的描述,可以通过组合这些“分面”来构造本体。“分面”应当有一个词表来描绘其领域内的现象,也应当有一套理论(关系、公理和约束等)来联系词表中的概念,并且这套理论应具有一致性。

图2建立了产品设计本体中有关设计的几个主要分面,如结构设计、工业设计,它们之间虽然有密切的联系,但也有相对独立的概念和理论体系。它们共享产品设计这一论域,是不同的概念模型。由此可见:本体的各个分面以子本体的形式构建,不仅提高了可用性,也具有较好的可重用性。

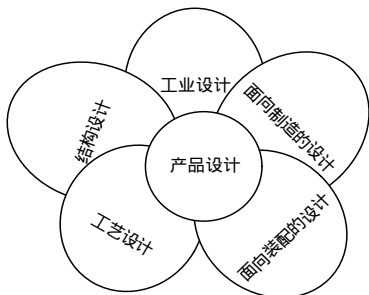


图2 汽车本体中与设计相关的几个主要分面

## 2.3 子本体的定义

早期本体层次化的含义主要是指顶层本体到应用本体的层次化,或者指本体内部结构由概念、实例、属性到属性值的层次化,考虑论域的层次化没有引起足够的重视。

标准上层本体(SUO)之一SUMO(如图3)就是一个模块化的本体<sup>[5]</sup>,每个方框代表一个子本体。

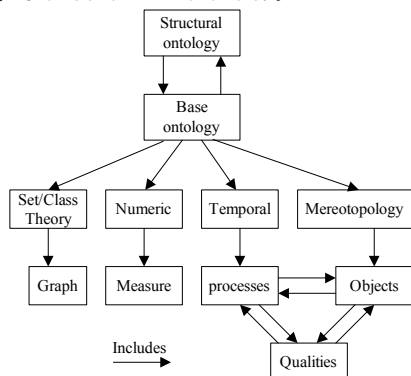


图3 子本体示例

统一以上模块化思想,把所有具有本体特性的组件统称为本体(或子本体),这样不仅可以简化本体的概念体系,更重要的是可以统一本体的逻辑语言。

**定义2(子本体)** 本体 $O_1$ 的论域是本体 $O_2$ 的论域的一个子集,则本体 $O_1$ 称为本体 $O_2$ 的子本体。

本体 $O_1$ 是本体 $O_2$ 的逻辑理论连续的子集。本体理论上应具有一定边界,实际中边界却是模糊的,因为世界是普遍联系的。本体的边界由论域确定。注意:子本体和其父本体之间的关系不同于子类 and 类之间的关系,不具备继承性。据此定义,上文中的包和视点都是子本体。

## 2.4 构建子本体的原则

早期的研究表明,越顶层的本体越容易重用,但越缺乏可用性。最近的研究表明,大型的顶层本体在重用时,也可能只需要它的一些局部,而不是全部,例如“单位”本体,在被“力学”本体重用时,就只需要力学领域必需的单位;底层的本体也有许多值得重用的部分。子本体将一个大型的本体分解为满足不同应用需求的一个个较小的模块,较好地解决了本体的可重用性和可用性这一对矛盾。

应当指出,适当的模块化可以增加本体的可理解性,从而降低本体构建的难度,同时也便于协同开发本体。但模块化却会增加知识的重复和冗余,增大本体存取开销,因此,模块化应该是受控的。构建子本体(模块化)的原则如下:(1)子本体是否有较清晰的边界和完整的语义,逻辑理论是否具有一致性。(2)是否有应用需求,即是否有可重用性和有用性。(3)其上级父本体是否太复杂,划分子本体是否可以简化本体的构造难度。

因此在创建子本体时,除了保证理论的完备性、一致性和简洁性外,还要充分考虑本体的可用性和可重用性的折中。例如,一个复杂产品一般都由机械系统、控制系统和液动(或气动)系统等子系统组成,它们分属不同的学科,具有不同的建模理论和工具。在构造复杂产品本体时,将这些子系统独立建模(如图4),有可能被一些简单产品直接重用,如机械产品重用机械系统本体。也可以将机械系统等继续分解为二级子本体,如动力系统、传动系统、执行系统。

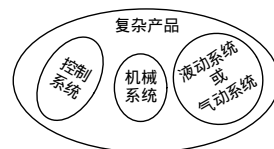


图4 复杂产品本体体系

## 3 面向本体操作的递归定义

通过以上分析,进一步给出原子本体和本体的递归定义。

**定义3(原子本体 $O_{atom}$ )** 符合定义1的基本特征,并且论域具有相对清晰的边界,具有一套用一阶语言描述的一致理论 $T()$ 和一个唯一的名字 $name()$ 。它是不能(即如果分解就会导致语义不完整)或不必要(分解后会产生太多的冗余或没有重用的价值)继续分解的本体。

这里“分解”是指将本体分割为2个和2个以上的子本体的过程。原子本体总是处于本体层次树的叶结点。

图5是笔者开发的基于本体的协同产品开发环境时用扩展的UML语言创建的“消息”本体。对于这个简单的本体,如果去掉其中的任何一个概念,都将使得本体失去完整的语义。因而这个消息本体是原子本体。

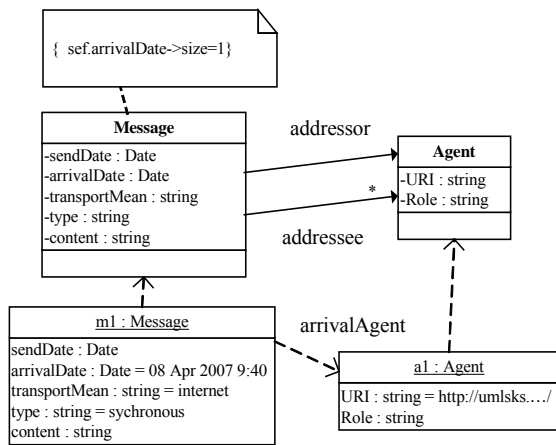


图5 用扩展的UML语言创建的消息本体

定义4(本体的递归定义)

$$O = \begin{cases} O_{atom} \\ O_i \cdot O_j \end{cases}$$

其中,  $\cdot$ 表示定义在本体上的二元算子,如融合、对齐、组合、集成、求交;  $O_i$ 和 $O_j$ 是原子本体或本体。

通过下例可以说明本体递归定义的意义。假设在创建木材本体时,考虑到主要的应用是装修业和造纸业。装修关心木材的强度、花纹、手感和香型等宏观特性;造纸关心木材的纤维长度、强度、色质和柔性等微观特性。这样就可以针对适合于装修的木材和适合于造纸的木材分别创建2个木材的子本体。必要时这2个本体可以经过“集成”形成木材本体(如图6)。

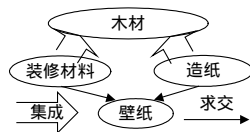


图6 本体模块化的必要性和方法举例

通过概念和应用分析不难发现:装修材料和造纸是具有交叉关系的2个本体。“壁纸”是装修的一种材料也是纸的一类,壁纸是装修本体和造纸本体的“交”。这样就可以十分方便地采用自下而上和自上而下相结合的方法(综合法)来建立本体。

(上接第47页)

[5] Witkin A P. Scale-space Filtering[C]//Proc. of the 8th Int'l Joint Conf. on Artificial Intelligence. Karlsruhe, Germany: [s. n.], 1983.  
 [6] Weickert J. Anisotropic Diffusion in Image Processing[M]. Stuttgart, Germany: Teubner-Verlag, 1999.  
 [7] Osher S, Sethian J. Fronts Propagating with Curvature Dependent

(上接第51页)

### 3 结束语

本文从工程化的视角出发,以平台设计的思想审视整个质检过程,抽象出质检任务、质检方案、质检模板、质检规则四要素构成质检模型,层次清晰,即能满足规则的灵活配置和扩展,又能针对不同的数据做到全面自动或抽样检查;作为一个质检软件平台,除了质检模块以外,还有标准构建模块、数据交换模块、错误查看模块、统计评价模块共5个部分组成,它们层层关联,组成一个有机的整体。

本设计的不足之处在于,规则的配置必须在抽象的接口

### 4 结束语

本体概念是本体工程的基石,对本体概念的深刻内涵和归一化处理的研究可以使本体工程沿着更科学的方向快速发展。在构建基于本体的知识库和基于本体的协同产品开发环境时,提出了自己的见解(定义1~定义4),这些新定义的意义在于:

(1)定义1明确地提出,任何一个本体都对应于一个论域,这与逻辑理论相一致。本体就是在这个论域上建立的知识模型,它构成一个(逻辑)理论。

(2)定义2用子本体统一了不规范的模块化思想。子本体和父本体之间的关系体现在论域的包含关系上,具有严格的数学意义。

(3)定义3和定义4说明可以对已有的本体通过一定的操作(运算)生成新的本体。

(4)这些本体定义使本体从无形趋于有形,从复杂变得简单;从模糊的大一统到清晰的层次化。

在此基础上,应当重新审视本体的分类体系,对现有本体的逻辑表示语言进行扩展和统一并进一步研究构成原子本体的充要条件和本体运算的具体而精确的定义,这样才能在真正意义上实现本体创建、进化、操作和维护的自动化或半自动化。

### 参考文献

[1] Bao Jie, Doina C, Vasant H. Towards Collaborative Environments for Ontology Construction and Sharing[C]//Proceedings of the International Symposium on Collaborative Technologies and Systems. Washington D. C., USA: IEEE Computer Society, 2006.  
 [2] Batres R, West M. An Upper Ontology Based on ISO 15926[J]. Computers and Chemical Engineering, 2007, 31(5/6): 519-534.  
 [3] Noy N F, McGuinness D L. Ontology Development 101: A Guide to Creating Your First Ontology[DB/OL]. (2007-06-30). [http://ontoworld.org/wiki/Ontology\\_Engineering#OTK\\_methodology](http://ontoworld.org/wiki/Ontology_Engineering#OTK_methodology).  
 [4] Takeda H, Iino K. Ontology-supported Agent Communication[DB/OL]. (2007-06-30). <http://www-kasm.nii.ac.jp/papers/takeda/95/aaai-syposium.pdf>.  
 [5] Gómez-Pérez A, Fernández-López M, Corcho O. Ontological Engineering[M]. London, UK: Springer Press, 2002.

Speed, Algorithms Based on the Hamilton-Jacobi Formulation[J]. J. of Comp. Physics., 1988, 79(1): 12-49.

[8] 严家斌, 刘贵忠. 一种新的各向异性扩散时间尺度估计方法[J]. 计算机工程, 2008: 34(2): 43-45.

上进行且要对业务有准确的理解,才能配置出适合业务需求的规则。同时 Geodatabase 数据模型的拓扑关系要根据需要临时构建,这需要消耗一定的系统资源和时间。

### 参考文献

[1] 吴秀芹. ArcGIS 9 地理信息系统应用与实践[M]. 北京: 清华大学出版社, 2002.  
 [2] 全国地理信息标准化技术委员会. 地理信息国家标准手册[M]. 北京: 中国标准出版社, 2005.  
 [3] Zeiler M. 为我们的世界建模: ESRI 地理数据库设计指南[M]. 北京: 人民邮电出版社, 2004.

