

●宋文,张剑,邵燕(中国科学院文献情报中心,北京 100080)

# 顶层本体研究

[关键词] 知识组织;本体;顶层本体

[摘要] 顶层本体是各领域本体的基础,对顶层本体进行研究具有非常现实的意义。本文从顶层本体研究的意义、概念、选择以及相关研究项目四个方面,对顶层本体进行介绍,并进行了简要分析,以便为国内本体的开发提供一些参考。

[中图分类号] TP18

[文献标识码] A

[文章编号] 1005-8214(2006)01-0043-03

最近几年,本体的研究与建设已成为语义网研究中的热点。本体包括顶层本体、领域本体、通用本体和任务本体,对国外顶层本体建设状况进行跟踪研究,对建设中文环境的本体数据库体系将产生积极的影响。

## 1 顶层本体研究的意义

顶层本体是独立于领域的高层本体,反映世界的普遍概念和常识。我们认为对顶层本体进行研究具有如下意义和作用:

- 本体是近几年国内外知识组织体系领域的研究热点,是实现万维网按照用户的问题进行语义检索和自动推理的基础设施,而顶层本体是各领域本体的基础。

- 充分利用国际本体研究成果,在现有的顶层本体基础上构建领域本体,直接引用顶层本体的概念、公理、断言,保持领域本体概念与顶层本体概念的一致性。

- 国内有不少传统的知识组织体系,如中国图书馆图书分类法、中国科学院图书馆图书分类法、汉语主题词表,以及许多的专业词表等,这些传统的知识组织体系是否可以按照国际主流的网络化知识组织体系构建模式去发展,并与国际标准的或通用的知识组织实行互操作。研究顶层本体,是将传统知识组织体系与国内外各领域本体进行互操作的第一步。

- 通过对顶层本体的研究,对国际上已开发的或正在开发的各种顶层本体做出判断和评价,以利于选择具备发展前景、在国际范围应用广泛的顶层本体,便于通过这样顶层本体的使用,实现我们自己构建的知识组织体系与其他的领域本体的互操作。

## 2 顶层本体的概念

### 2.1 顶层本体的定义

顶层本体是高层的独立于领域的本体,它使完全相异的系统可以使用一个共同的知识库,并且从顶层本体可以衍生出领域本体。顶层本体通常表达常识性概念,也就是说是人类对世界的基本认识。顶层本体的概念主要局限在基本的、普遍的、抽象的和哲学上的概念。标准顶层本体也称为基本本体或通用本体。

### 2.2 顶层本体与中层本体

中层本体在顶层本体的抽象概念与领域本体中领域专指概念之间搭起一座桥梁。从理论上说,本体之间是可以映射的,而中层本体和顶层本体使领域本体之间的映射更加容易。中层本体使顶层本体中的抽象概念表达得更加具体。这个层次的本体可以有許多类型,一般表达普遍使用的概念,如时间本体、地点本体等,这些本体有时也称作实用本体。

### 2.3 中层本体与领域本体

领域本体表达领域特有的概念及其概念之间的关系。如果同一个概念出现在多个领域中,由于不同领域的关系,概念的表达可以有很大不同。领域本体可以在中层本体内建设,领域本体的概念可以由中层和顶层本体的概念扩展而来。重用已建成的本体能够使我们利用被重用本体的丰富的概念和逻辑语义关系。顶层本体的重用是为了领域本体的概念可以从顶层本体产生或将领域本体的概念映射到顶层本体的概念。用这个方法,本体可以提供概念分解的语义网络。使用共同的中层和顶层本体可以使领域本体的集成和映射更加容易。

## 3 顶层本体的作用

万维网为人们提供了大量丰富的信息,语义网是万维网的进化,它试图准确地捕获数据的含义,使得软件能够解释它们。语义网的关键成分是定义概念和概念之间关系的本体。因为本体能提供数据的关系,所以搜索引擎能根据信息的情景关系按照概念的含义执行语义检索,而不仅仅是按照句法的文本字符串检索。例如,如果一辆汽车和一匹马都叫“mustang”,用语义检索方法,就可以将它们区别开来。本体丰富的语义关系使系统能够根据一些简单的事实从而推导出新的事实,能够推导出隐含在本体中的一般性知识。当数据和应用映射到本体后,推理引擎能够改进对数据的理解和发现,以及对 Web 服务等应用的发现和组合能力。顶层本体定义了用于中层和领域本体的基础概念。如果领域本体的概念是从标准顶层本体继承过来的,那么这些领域本体之间的映射就会更容易。顶层本体可以有多方面的用途,主要用途概括为:

- 用于设计新的领域本体。领域本体的设计者可以用公用的顶层本体定义概念,建设新的本体,由此获得与相关系统互操作的能力。

- 已有的本体的重用/集成。现存本体中的概念可以映射到顶层本体,当相关的领域本体都映射到顶层本体后,可以方便地实现领域本体的集成。

- 领域本体的互操作。相关的领域本体都映射到公用的顶层本体后,这些本体之间因为共享术语和定义,所以可以实现互操作。

可以有两种使用顶层本体的方法:自顶向下法和自底

向上法。自顶向下法是从顶层本体获取领域本体的概念,用这个方法可以利用顶层本体已有的知识和经验,同时顶层本体也为领域本体的构建提供框架。自底向上法是本体的设计者将领域本体映射到顶层本体。这种方法也利用了顶层本体的知识,但这种映射将会有许多困难,因为领域本体和顶层本体概念之间存在不一致性。

#### 4 顶层本体的选择

顶层本体是客观世界的反映,在本体构建中面临很多的选择,这些选择来源于人们对世界的认识。不同的人对客观世界的事物、现象有不同的认识和理解,由此,不同设计者建设的顶层本体在概念体系结构上会有很大的差异。本体的选择是评价本体的重要标准。本体的选择包括描述的和修正的,繁杂的和简化的,通用的、个别的和集合的,连续的和断续的,三维的和四维的,部分的和整体的等等。其他选择包括对属性的构成有哪些,对变化应该如何表示的不同认识,还有对粒度、模糊等的不同理解。其中的很多选择之间联系都是错综复杂的,比如讨论“连续”和“断续”就涉及到“三维”和“四维”。由于篇幅有限,笔者只做一些最基本的介绍。

##### 4.1 描述的与修正的

从本体建模角度出发,可以将本体分为描述的和修正的本体。比如,一个人是怎样把这个世界概念化的,以及本体工程的产品是什么或应该是什么。描述性本体反映的是人们对世界的认识,修正性本体力图反映世界的本来面目。一个描述性本体力图从自然语言和人类认知来获得更具常识性和社会性的概念,更重视认识世界的中间媒介(如语言等),而不太重视从科学和哲学的角度出发去考虑。而一个修正性(有时也称说明性)本体则着重于从科学和哲学的视角出发,将它们的选择构建在科学理论和哲学视角,努力认识世界的本来面目,所以修正本体是说明这个世界的。修正本体不按照某个历史阶段对世界的认识去构建本体,因此,修正性本体声称它们构建的模型是关于世界本来面目的。例如,在描述性本体看汽车和修车,车是一个抽象出来的独立于时间的对象类,而修车是一个必须有时间属性的对象类,比如,有开始和结束的时间。而修正性本体认为汽车和汽车修理都是具有时间-空间性的对象(事物-事件)。

##### 4.2 繁杂的与简约的

一个繁杂的顶层本体尽情地用概念来表示现实世界中的事物,繁杂本体的概念数量是很庞大的。与此相反,简约的本体描述最原始的概念,用最少的“元概念”来产生复杂概念。

在 WonderWeb 基础本体库中,语言学和认知工程的描述性本体 DOLCE 和基本正式本体 BFO 都是繁杂和描述性的,而以对象为核心的高层参考本体 OCHLRO (Object-Centered High-Level Reference Ontology) 则是简约和修正的。

在本体的选择(描述性的和修正的,繁杂的和简约的,等等)中,许多选择的决定其实都是有一些假设前提的:如何看待这个世界(比如,严格的现实主义,不带丝毫不同的可能性),以及这个世界或部分世界的设计模型可以达到什么样子。因此,许多本体选择都会同时存在。比如,修正的和简约的一般会在一起。

#### 4.3 三维与四维

3D (3-Dimension) 和 4D (4-Dimension) 常常与持续 (Endurants) 和断续 (Perdurants) 等同起来。3D 和 4D 是本体的一个基本选择,是关于对变化的认识。一个实体的变化意味着什么?

一般 3D 观点认为物体是:①在一个三维空间中延伸;②在它们生命的每一刻都完全展示;③变化的实体在不同的时间具有不同的属性。比如,一个人可以说当我在阳台上时,我的手比现在还冷。相反的是,一个四维的观点认为物体是:①时空的蠕虫;②在每一刻都只是部分的展示;③不同的阶段变化的实体会具有不同的属性,比如,在阳台的这段时间里我的手比现在还更冷。

#### 5 顶层本体项目

许多在研的顶层本体项目中,有两个比较重要的项目是 IEEE 标准顶层本体工作组项目和 Wonder Web 的顶层本体项目。

##### 5.1 IEEE SUO 工作组

IEEE 标准顶层本体工作组由 IEEE 计算机协会标准工作委员会赞助,IEEE 标准化协会管理。其任务是选择一个顶层本体作为标准顶层本体,这个本体能够使计算机进行诸如数据互操作、信息检索、机器推理和自然语言处理等工作。IEEE 标准顶层本体项目的目标是:

(1) 自动推理。标准顶层本体应能进行自动逻辑推理来支持基于知识的推理应用。

(2) 互操作。标准顶层本体将为实现不同软件和数据库应用之间的互操作性提供一个基础。

(3) 应用。应用在需要在语义层上实现互操作的各种领域的电子商务系统;提供学习者直接从公用的标准顶层本体中学习概念和他们之间的关系;协助进行自然语言处理,消除自然语言表达上的歧异。

IEEE 标准顶层本体工作组目前有五个候选顶层本体,分别是:SUMO (Suggested Upper Merged Ontology)、OpenCyc、IFF (Information Flow Framework)、4D Ontology 和 MSO (The Multi-Source Ontology of WebKB-2)。

##### 5.1.1 IFF

IFF (Information Flow Framework) 是 IEEE SUO 工作组正在研发的标准顶层本体,由 Bob Kent 主持开发。IFF 主要是一个元本性,创建者用范畴理论和信息流理论支持本体的模块化构造。IFF 为构造本体、映射本体和集成本体提供框架。

##### 5.1.2 4D 本体

顶层本体的设计思想实际上反映了设计者对世界的认识和看法,在顶层本体建设中有 3D 和 4D 两种观点,主要是对对象或实体变化的认识。IEEE SUO 工作组在 2003 年 10 月正式通过了 4D 本体 (4-Dimensional Ontology) 的工作方案。其目标是按照 4D 的观点组织本体的概念,并使该本体成为 IFF 的一个组成部分,并建设与 IFF 框架中的其他本体的映射。

##### 5.1.3 SUMO

SUMO (Suggested Upper Merged Ontology) 最初由 Ian Niles 和 Adam Pease 开发,现在由 Teknowledge Corporation 维护。SUMO 是 IEEE 标准顶层本体工作组考虑的三个最初文档之一。SUMO 的建设目标是实施数据的

互操作、信息检索、自动推理和自然语言处理。SUMO 包括人类认知方面的类别和现实描述的类别。SUMO 合并现有的顶层本体而成,被合并的本体包括:Ontolingua 服务器上可获得的本体、John Sowa 开发的顶层本体、ITBM-CNR 开发的顶层本体、Russell 和 Norvig 开发的顶层本体和各种拓扑理论。SUMO 的知识表示语言是 SUO-KIF 语言(美国国家标准 KIF 格式的简化版),可以按 GNU 公共许可协议使用。目前, SUMO 有 1770 个概念, 7278 个断言, 1240 个规则。SUMO 设计者主张模块化设计本体, 将本体分成多个子本体。最高层概念是“实体(Entity)”, “实体”又进一步分为(Physical)物理实体和(Abstract)抽象实体。物理实体又分为对象和过程, 抽象实体包括集合理论、属性、关系、数量和度量、时间概念, 如时期以及部分和整体。

#### 5. 1. 4 OpenCyc

OpenCyc 是 Cyc 技术的开放版本, Cyc 的创建公司 Cycorp 建立了独立的机构 OpenCyc.org 来发布和管理 OpenCyc。通过这样的管理方法, 建立了一条将 Cyc 技术从 ResearchCyc 到 OpenCyc 的通道。OpenCyc 作为常识知识库用于支持自然语言处理。OpenCyc 用 CycL 语言表示, CycL 与 KIF 非常类似, 与 KIF 遵循相同的命名规则。OpenCyc 本体有 10 万条原术语, 6000 个概念, 6 万条公理。Cyc 是模块化构造的, 被分成多个微理论(即模块), 每个微理论包含一组断言和假设。一个微理论可以由遗传或扩充另一个微理论而成。OpenCyc 中最高层实体是“Thing”, “Thing”又分成“Individual”“PartiallyIntangible”和“MathematicalOrComputationalThing”。

#### 5. 1. 5 MSO

WebKB-2 是一个知识服务器, MSO (The Multi-Source Ontology) 是 WebKB-2 上的多来源本体, 该本体是各种顶层本体和词汇本体(来源于 WordNet)的集成。在 MSO 中, 每一个对象(类、关系、陈述)都保留了其来源, 除了词汇本体, 所有来源本体的语义也都全部被保留下来。因为各种本体的类别和类别之间的连接都完好地保存下来, 所以各种本体的类别和陈述可以互相补充、说明, 可以避免词汇冲突(当增加新的类别或陈述时可以察觉语义冲突)。相对于传统本体来说, 这个知识库可以更好地保持内部连接完好。而传统的本体更趋向独立开发, 如果想要重用本体就需要经过选择、合并等步骤。

MSO 也许是本体创建者将其本体与其他本体互联的一种有效手段, 它可以供人们查询、重用或补充本体。最重要的是 MSO 的创建不依靠中央规范控制, 概念间的冲突不需要通过讨论和妥协来解决。

SUO 工作组于 2004 年 5 月通过投票决定 MSO 作为 IEEE 标准顶层工作组的候选文档。

#### 5. 2 WonderWeb

WonderWeb 是欧盟信息社会委员会的研究项目, 项目联盟成员包括大学和工业界, 项目同时与 DARPA DAML 项目和 W3C 合作。项目的目标是建设支持语义网的本体基础设施。项目的主要目标包括: 本体语言的研究和制定, 本体构建模式的研究, 基础本体库的建设以及应用工具的建设等。

与 IEEE SUO 工作组不同的是, WonderWeb 并没有

打算建设一个通用的标准顶层本体, 其目标是用本体工程的方式模块化建设基础本体库, 这个基础本体库可以覆盖广泛的应用领域, 能作为领域本体建设的基础。目前基础本体库已有三个模块: DOLCE、OCHRE 和 BFO。WonderWeb 项目对顶层本体建设的另一贡献是系统地分析和阐述了在本体设计和构建中的一些基本思想和原则性的选择。如通用和特殊、具体和抽象、持续和断续等, 成为评价顶层本体的理论, 也为领域本体建设提供参考和借鉴。

#### [参考文献]

- [1] IEEE SUO. Overview of the SUMO [EB/OL]. <http://ontology.tekknowledge.com/arch.htm>, 2004-12-20.
- [2] Phytilla, Chris. An Analysis of the SUMO and Description in Unified Modeling Language [EB/OL]. <http://ontology.tekknowledge.com/Phytilla/Phytilla-SUMO.html>, 2004-12-25.
- [3] Ian Niles, Adam Pease. Towards a Standard Upper Ontology [EB/OL]. <http://home.earthlink.net/~adampease/professional/FOIS.pdf>, 2004-12-31.
- [4] Sevchenko, M. Online Presentation of an Upper Ontology [EB/OL]. <http://ontology.tekknowledge.com/Sevchenko.pdf>, 2004-01-20.
- [5] Iniles. SUMO (Suggested Upper Merged Ontology). Merge.txt. Revision [EB/OL]. <http://ontology.tekknowledge.com/cgi-bin/cvsweb.cgi/SUO/Merge.txt?rev=1.60&content-type=text/xcvsweb-markup>, 2005-01-15.
- [6] IEEE SUO. Suggested Upper Merged Ontology (SUMO) [EB/OL]. <http://www.ontologyportal.org>, 2004-11-28.
- [7] Salim K Semy, Mary K Pulvermacher, Leo J. Toward the Use of an Upper Ontology for U. S. Government and U. S. Military Domains [EB/OL]. <http://colab.cim3.net/file/work/SICOP/resources/UpperOntologyUsealong.doc>, 2005-02-20.
- [8] OpenCyc.org. OpenCyc: Project [EB/OL]. <http://www.opencyc.org/>, 2005-02-25.
- [9] Information Society Technologies [EB/OL]. <http://www.cordis.lu/ist/home.html>, 2005-05-11.
- [10] Claudio Masolo, etc. Wonder Web Deliverable D18: Ontology [EB/OL]. <http://wonderweb.semanticweb.org/deliverables/documents/D18.pdf>, 2005-03-13.
- [11] 邓志鸿, 等. Ontology 的研究综述 [J]. 北京大学学报, 2002, (9): 728-730.

[作者简介] 宋文(1961-), 女, 中国科学院文献情报中心研究员, 发表论文多篇; 张剑(1981-), 女, 中国科学院文献情报中心 2003 级研究生; 邵燕(1981-), 女, 中国科学院文献情报中心 2003 级研究生。

[收稿日期] 2005-05-21 [责任编辑] 肖群