

语义网中重量级本体的设计

刘 康 黄 奇

南京大学信息管理系 南京 210093

[摘要] 通过对本体论中顶层本体、领域本体和任务本体的分析,结合语义网的应用实际,提出语义网中重量级本体设计的基本概念、基本要求和基本关系;认为角色概念有丰富的语义描述能力,是重量级本体区别于轻量级本体的重要特征,并试图通过对角色概念的剖析,说明角色概念对语义约束的描述特征。

[关键词] 语义网 本体 重量级 设计

[分类号] G250.76;G354.4

Methodology of High-weight Ontology Designing for Semantic Web

Liu Kang Huang Qi

Department of Information Management, Nanjing University, Nanjing 210093

[Abstract] Ontology is critical to semantic Web, the ability of knowledge expressing depends on the way of ontology designing. This paper introduces the definition and category of ontology, points out the characteristics of ontology when using semantic Web. Then it introduces the heavy-weight ontology which differentiates from metadata. In addition, it also discusses the methodology of designing heavy-weight ontology, especially the domain ontology. Finally, the paper discusses the implement language of ontology OWL.

[Keywords] semantic Web high-weight ontology designing methodology

1 语义网及本体简介

语义网是为了解决当前基于HTML的Web的缺陷而提出的。由于HTML语言的设计局限性,使得当前的Web关注于数据和信息的显示,而忽略了数据和信息之间的关系,导致Web自动化受到了很大的阻碍。语义网是当前WWW的扩展和延伸,随着互联网的发展及其应用的深化,以HTML为核心的Web技术已经无法满足电子商务、电子政务、e-Learning、e-Science发展的需求。语义网为解决这些问题提供了广阔的前景,它通过对信息的充分定义,使得计算机和人能够更好地协作^[1]。

作为互联网上信息组织的全新的形式,语义网已经成为信息资源管理的研究前沿。

语义网的理论模型是一个分层的七层架构,其中,本体词汇层(ontology vocabulary)扮演着关键性的角色,它决定了Web元素含有语义的能力、语义的正确性和推理的能力。为此,W3C于2004年2月公布了描述本体的语言标准OWL(本体Web语言)及其三个子集OWL-Full、OWL-DL和OWL-Lite。

计算机界的本体研究始于20世纪90年代初期的知识基础社区(knowledge base community)研究^[2],其快速发展却得益于语义Web研究的强大势头。本体在不同的研究领域有不

同的定义。

1.1 本体的定义

目前常见的本体定义有如下几种:

1.1.1 哲学定义 本体论即有关“存在”的理论。本体是为了解释存在的事物是什么,通过引入一个临界的分类系统来构成物质世界,并描述事物和它们的本质联系的理论。本体是客观存在的一个系统的解释和说明,关心的是客观事物存在的抽象本质。顶层本体是哲学上本体研究的一个重要问题。

1.1.2 人工智能定义 本体乃是概念模型的明确的规范说明^[3]。从基于知识的系统的观点来看,本体的定义是:用来作为组建信息处理系统的元素的概念的理论。本体是表示领域中的实体、类、属性、断言和功能特定词汇以及这些词汇之间的关系的集合。

1.2 语义网上的本体

在语义网的设计中,本体的作用是为领域提供可共享的理解。因此语义网上的本体可参照人工智能中的本体概念。根据Web上问题的特点,语义网上使用的主要是领域本体和任务本体。

根据“概念模型的明确的规范说明”这一定义,可知本体设计有三个特点:conceptual(概念化的)、explicit(明确的)、formal(形式化、规范的、可为计算机所理解的)。

2 本体的分类

现有的本体分类体系有三种:①根据本体所研究的问题的领域,可将本体分为顶层本体和非顶层本体;②根据本体处理问题的类型,可将本体分为领域本体和任务本体;③根据本体描述语义的能力,可将本体分为轻量级本体和重量级本体^[4]。

2.1 顶层本体

顶层本体(upper-level ontology)是哲学上本体研究的重要课题,早在2000年前亚里士多德就开始了这方面的探索。顶层本体尝试将客观世界的一切存在事物用高层次的类目加以区分,以达到解释客观世界的目的。亚里士多德给客观世界做了10种分类,分别是物质、数量、质量、关系、位置、时间等。此后C. S. Peirce又提出了第一性(firstness)、第二性(secondness)、第三性(thirdness)名词的区分^[5]。其中,第一性名词是无须参考其他名词而直接可以定义的名词,如人类、铁等;第二性名词是必须参考其他事物而定义的名词,如朋友、教师等;第三性名词是提供第二性名词存在的环境(上下文)的词汇,比如友谊、学校等。在人工智能领域,J. Sowa引入了第三性名词,对人工智能研究产生了巨大的影响。

2.2 领域本体和任务本体

对于解决实际问题的本体,根据其解决问题的类型,可以将其分为领域本体和任务本体两类。在实际的问题求解过程中,通常先使用领域本体描述系统,接着通过任务本体进行求解或者完成任务。领域本体是用来描述执行任务的领域的知识,例如自行车本体、学校本体;任务本体是描述基于知识的任务处理系统的计算性结构,例如诊断本体、监控本体。

2.3 重量级本体(heavy-weight ontology)和轻量级本体(light-weight ontology)

根据本体的语义丰富程度,可将本体分为重量级本体和轻量级本体。其中,重量级本体对概念的含义有严格的要求,从哲学的观点来组织概念,概念之间有精确的语义联系。顶层本体就是一个典型的重量级本体。而轻量级本体则无需严格地考虑概念的定义和组织,它主要是用来建立一个概念的层次关系,为某些应用服务,例如yahoo的搜索分类目录。

元数据也是一种轻量级本体。以RDF为代表的元数据技术,在语义网中起着描述资源的作用。但是元数据的语义能力不足以满足建立语义网的语义需求。本体词汇层的设计,就是为了更多地容纳重量级本体,提供强大的语义为上层的高级应用服务。因此,如何设计出良好的、可用的、语义丰富的重量级本体,对于语义网的建构至关重要。

3 语义网上本体的设计

在领域本体和任务本体的设计中,类、属性、关系、角

色是重要概念。只有正确地认识和使用这些概念,才能设计出语义正确的良构的本体。领域专家对领域的准确认识也是建设良好领域本体和任务本体的关键。

3.1 类的定义

类的定义对于语义的明确性非常重要,它需要用到一个“固有属性”的概念。

3.1.1 定义一 固有属性即事物之所以存在的一个固有的属性。若某事物丧失了它的固有属性,则它便不能成为这样一个事物。

例如在本体中,“动物”这个类具有“可以活动”这样一个固有属性。固有属性是建立类的重要标准,是本体中的类和面向对象中的类的定义的一个重要区别,也是确保本体语义正确性的一个重要准绳。

子类和超类在本体中可以用is-a关系来表达,is-a关系的确定也是本体建设中的一个难点。在重量级本体的设计中,is-a关系的正确使用直接关系到语义的正确性,对于基于本体的推理的成败至关重要。

3.1.2 定义二 只有一个类A的所有实例(instance)都是另一个类B的实例的时候,才有类A是类B的子类(A is-a B)成立。据此,可知如下一些is-a关系是不恰当的:

如果认为<Teacher is-a human>,则<Mr. A instance-of teacher>是不成立的,因为如果某人A不再当老师,即A不是教师类的实例了,根据定义二,则推出A这个实例不再是人类的实例,可见语义便产生了错误。因此在本体中不允许以上的定义。这个特点,是本体中类的概念和面向对象中类的概念的一个重要区别。

事实上,可以通过引入“角色”概念来解决此类问题,即认为“教师”是一个角色,由某个类的实例来扮演这样一个“角色”,这种方法是本体论模型中的流派之一。

此外,认为<物体 is-a 一定量的物质>也是不正确的:物质的分割不会影响其固有属性,物体是由一些物质所构成(如木桌是由一定量的木头构成),但物体的分割会使得其固有属性发生变化。

3.2 部分整体关系和一般关系

部分整体关系连接概念和部分概念,一般关系表达整体概念中各部分概念的关系。

部分整体关系表达的是整体概念和部分概念之间的关系。如图1所示,“Institute of education”这个整体概念就拥有如“Teacher”、“Student”等部分概念;各个部分之间的关系则用一般关系来表达,如部分概念“Age”和“Min Teacher Age”之间的数量大小关系(“Larger than equal”)就是一种一般关系。

需注意的是,部分整体关系是不可传递性的,否则会出现<手 part-of 人> & <人 part-of 学校>推理得出<手 part-of 学校>的错误结论。

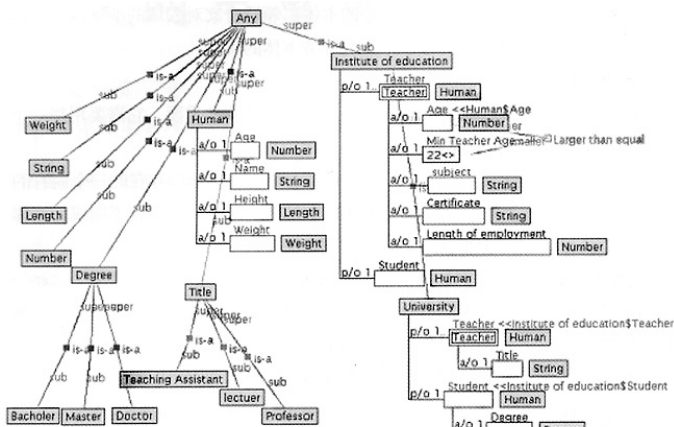


图1 本体片断

一般关系是用于表达各部分概念之间的各种关系,比如数量关系、空间位置关系等。不同概念的部分概念之间不能建立一般关系。

3.3 角色^[6]

不同学术观点的本体学者,对于角色概念在本体中的重要性有不同的认识。笔者认为角色概念丰富了本体的语义,对于本体的语义正确性保证也有不小的作用。根据Mizoguchi教授的观点,角色是基本概念构成的复杂概念所派生的。比如对于一个人来说,当他具有一个职业诸如教师的时候,他便扮演了“教师”这样一个角色。但是教师并非所有人都可以担当,可能需要某些条件。这样就产生了“基本概念”、“部分-角色概念”和“角色持有者”概念,如表1所示:

表1 部分-角色概念、基本概念和角色持有者

基本概念	部分-角色概念	角色持有者
人	教师角色	教师
[B1]名字	[R1]名字	[R1]名字
[B1]年龄	[R2-1]年龄(>22)	[R2-1]年龄(>22)
[B2]身高		[B2]身高
[B2]体重		[B2]体重
	[R2-2]教授科目	[R2-2]教授科目
	[R2-2]工作年限	[R2-2]工作年限
	[R2-2]拥有证书	[R2-2]拥有证书

当一个基本概念(如“人”)在某个上下文关系(如“学校”)中扮演某个角色(如“教师”)的时候,这个基本概念的某些属性(如“年龄”)就会有某些语义约束(semantic constraint,如教师年龄必须大于22岁);同时作为角色持有者(如担当教师角色的某个人),又会具有某些基本概念中不具备的新的属性(如执教年限、所教课程等)。因此,“基本概念”、“角色概念”和“角色持有者”概念之间,就会有如图2所示的一个映射关系。

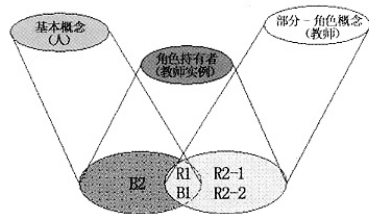


图2 基本概念、部分-角色概念和角色持有者的映射关系

由图2可知,基本概念充当角色概念的时候,它的属性会被角色概念有选择地引用,因而角色概念中既包含有基本概念中的一些根本属性(如R1、B1),又具备反映其自身特点的新属性(如R2-2)以及对基本概念属性的进一步规定(如R2-1)。基本概念

中不符合部分-角色概念特点的属性不被引用(如B2)。因此,充当角色概念的实例(角色持有者),拥有了基本概念和角色概念的所有属性。

3.4 继承

在本体中,子类从父类继承属性、部分-整体关系,同时可以表达特别的继承。如图1所示,university可以继承institute of education(教育机构,代表各种形式的学校)的teacher、student概念,同时可以特别地规定teacher具有title属性、student具有degree属性。

通过以上的规则,我们可以大致建立本体的模型,图1所表示的就是一个学校本体的片段。它表达了学校概念的一些要素以及作为学校类的子类的大学的要素,还有各个部分概念之间的关系,等等。这和分类系统、以元数据为代表的轻量级本体是有着明显区别的。

4 用OWL表达本体

OWL由W3C组织于2004年颁布,它由美国的DAML(DARPA代理标记语言)和欧洲的OIL(本体交换语言)两大本体语言标准演化而来,并获得了各界的一致认可。OWL是未来Web上本体语言的标准。

为了兼顾描述能力和推理能力,OWL设有三个种类,分别为OWL-full、OWL-DL和OWL-Lite,描述能力方面OWL-full强于OWL-DL,OWL-DL强于OWL-Lite,而推理能力则恰好相反。据此,我们可以根据描述领域的推理及描述能力的需求选择OWL语言的种类。

三种OWL之间又具有这样的描述和推理关系^[7]:一个合法的OWL-lite本体必然是一个合法的OWL-DL本体,一个合法的OWL-DL本体必然是一个合法的OWL-full本体,一个正

确的OWL-Lite本体推出结论必然是一个合法的OWL-DL本体推出结论,一个正确的OWL-DL本体推出结论必然是一个合法的OWL-full本体推出结论。

根据图1中已经建设好的本体模型,我们可以写出OWL文件表达学校本体(片断):

```
.....
</owl:Ontology>
<owl:Class rdf:ID="WholenessConcept">
<rdfs:label>WholenessConcept</rdfs:label>
</owl:Class>
<owl:Class rdf:ID="RelationalConcept">
<rdfs:label>RelationalConcept</rdfs:label>
</owl:Class>
<owl:Class rdf:ID="Human">
<rdfs:label>Human</rdfs:label>
<rdfs:subClassOf rdf:resource="#Any">
<rdfs:comment>
.....
<owl:Class rdf:about="#Human"/></rdfs:domain>
<owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="Height">
<rdfs:domain>
<owl:Class rdf:about="#Human"/></rdfs:domain>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="Weight_role">
<rdfs:domain>
<owl:Class rdf:about="#Human"/></rdfs:domain>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="Teacher">
<rdfs:domain>
<owl:Class rdf:about="#School"/></rdfs:domain>
</owl:ObjectProperty>
```

```
<owl:ObjectProperty rdf:ID="larger">
<rdfs:domain>
<owl:Class rdf:about="#Larger than equal"/></rdfs:domain>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="smaller">
<rdfs:domain>
<owl:Class rdf:about="#Larger than equal"/></rdfs:domain>
</owl:ObjectProperty>
</rdf:RDF>
```

在OWL描述的本体的基础上,我们就可以使用XML解析器,通过程序实现本体的检索查询、基于本体的代理应用。同时还可以利用OWL-QL, OWL-S等工具,实现OWL查询和标识Web服务。

参考文献:

- 1 Tim Berners-Lee, James Hendler, Ora Lassila. The Semantic Web. Scientific American, 2001(5):28-37
- 2 Mizoguchi R. Tutorial on ontological engineering. New Generation Computing, 2004,22(2):8-9
- 3 Tom Gruber. What is an ontology?[2005-10-25].<http://www-ksl.stanford.edu/kst/hat-is-an-ontology.html>
- 4 Mizoguchi R, Johan Vanwelkenhuysen, Mitsuru Ikeda. Task Ontology for Reuse of Problem Solving Knowledge. In: Mars N J I. Towards Very Large Knowledge Bases. Virginia: IOS Press, 1995.46-59
- 5 Sowa J. Distinction Combination, and Constraints. In: Proceedings of IJCAI-95 Workshop on Basic Ontological Issues in Knowledge Sharing, 1995.13-15
- 6 Kozaki K, Yoshinobu Kitamura, Mitsuru Ikeda. Development of an environment for building ontologies which is based on a fundamental consideration of "relationship" and "role". Journal of JSAI, 2002,17(3):196-208
- 7 Grigoris Antoniou, Frank van Harmelen. A Semantic Web Primer. Massachusetts :Massachusetts Institute of Technology, 2004.154-236

〔作者简介〕 刘 康,男,1982年生,硕士研究生,发表论文1篇。

黄 奇,男,1961年生,教授,国家信息资源管理南京研究基地主任,发表论文30余篇,主编(参编)著作7部。

百度、搜狐、卓越加入“Apabi图书营销联盟”

由全国400多家出版社、100多家书店、数十家网上书店和30多家网站组成的“Apabi图书营销联盟”于2006年5月16日在京成立。据了解,这是迄今为止,我国规模最大的图书营销机构。

百度、搜狐、卓越网等诸多互联网知名企业,在“2006中国电子书产业年会上”,与方正Apabi、国家图书馆、商务印书馆、浙江省新华书店等龙头企业结成战略联盟,共同打

造立体化图书信息网络,旨在为读者提供便捷准确的找书、买书服务。

Apabi图书营销联盟致力于图书信息的共享与互通,充分利用互联网技术,将全面、准确、及时的图书信息以最便捷的方式展现给读者,方便读者随时购买、阅读图书,促进图书产业和互联网产业的融合和共同发展。

(本刊记者)