

一种网络环境中的本体演化和维护模型^{*}

王 进 陈恩红 林 乐

(中国科学技术大学计算机系 合肥230027)

An Ontology Evolution and Maintenance Model in Web Environment

WANG Jin CHEN En-Hong LIN Le

(Department of Computer Science, University of Science and Technology of China, Hefei 230027)

Abstract With the rapid development of Internet, Knowledge Web will be one of the most important development directions of the next generation of Internet. The applications of ontology will bring new opportunities and challenges to it. In this paper, we propose an ontology evolution and maintenance model in the Web environment to solve the conflicts and inconsistencies between domain ontologies and knowledge caused by the changes of the knowledge sources in the Web environment by extracting and analyzing the meta-data of the data sources.

Keywords Ontology, Evolution, Maintenance, Web environment

1. 引言

信息时代即将跨入知识 Web 的今天,知识的机器可处理性已经成为未来网络发展的主要目标之一。随着语义 Web 概念的提出,网络信息的语义也将会成为机器进行数据处理的重要依据。对此有关的学者提出了以 XML/XML Schema 和 RDF/RDF Schema (Resource Description Framework Schema)作为语义互操作的载体^[1,2],在此基础上,融合了哲学中本体(Ontology)的思想,结合了具有形式语义和推理支持的描述逻辑、富含认知建模原语的基于框架系统以及基于 XML 和 RDF 语法的网络语言等内容^[3],提供了基本的建模原语以及建立本体的广泛共享模型、规范的语义描述,并有效地使用逻辑推理进行知识处理,以期达到语义解释的最佳效果^[4~7]。这一理念在知识网络领域得到了广泛的认可,并且在很多的具体应用领域如电子商务中使用本体充当通信的中介、知识管理中应用本体作为组织知识仓库的有效技术手段、多 agent 系统的数据交换和共享语义平台以及搜索引擎中的语义载体等都体现出了相当的优点和广阔的前景。

在网络环境中,散布在不同地点的不同种类知识源中包含了各种各样结构化、半结构化和无结构的信息,如何有效地对它们进行描述、管理和使用是我们面临的一大课题,作为特定领域知识的层次化描述,本体所包含的相关领域概念、概念之间的关系以及有关这些概念和关系的规则共同提供了一个合适的形式和一个基于共享的术语集来对知识源的内容进行描述。因此,本体成为网络环境下知识表示的一个理想选择,并结合自然语言处理、信息检索、数据库、电子商务、知识管理以及多 agent 系统等共同构建了新一代知识网络的应用基础,具有极大的发展潜力和研究价值。

正因如此,人们对本体及相关应用领域如 Semantic Web 投入了很大的精力,并取得了相当的成效,特别是 World Wide Web Consortium(W3C)给出了一系列的标准和规范,如 RDF 模型和语法规则、RDF 词汇描述语言 1.0 以及 2002 年

11 月份发布的 Web 本体语言(OWL)1.0 等。现阶段,在基于本体的应用研究方面有相当的进展,如基于本体的知识库表示、不同类型知识源的信息集成、语义 Web 上的信息检索技术和基于本体的知识管理系统等。然而,在动态领域本体的变化检测以及本体模型演化和传播的问题上,仍然没有能够提出一种较好的思路和方法来给予解决,从而在一定程度上限制了领域本体应用上的灵活性和时效性,需要较程度的人为方式来进行协调。

本文提出一种在网络环境中的本体演化和维护模型,试图解决领域知识动态变化情况下,通过对知识源的元数据抽取和分析来支持本体模型的演化和维护问题。本文第2节介绍了模型的主要理论基础,第3节给出了模型的框架,并就每个模块的基本功能和设计思路作了简要介绍;最后总结并对今后的工作进行展望。

2. 模型的相关概念

本体概念结合了具有形式语义和推理支持的描述逻辑、富含认知建模原语的基于框架系统以及基于 XML 和 RDF 语法的网络语言等三个方面的内容,提供了基本的建模原语以及建立本体的广泛共享模型,规范的语义描述,并有效地使用逻辑推理对知识进行处理,以达到语义解释的最佳效果。

2.1 本体的构建

本体定义的主体包括该本体里的其他模型的参考列表、应用于该本体的规则列表以及类定义(class-def)和槽定义(slot-def)。

本体的类定义包括五个部分,分别是定义的类型、类名、文档、类继承声明和槽限制。定义的类型分为派生的(primitive)和定义的(defined)两种,用来说明该定义是类实例的必要条件或者是充要条件,缺省为派生的;类名给出了类定义的名称;文档则是对该类的描述;类继承声明指出了该类所继承的父类表达式列表;槽限制是一种特别的类表达式,所定义的类型也是每个槽限制的子类。

^{*} 本文研究得到国家自然科学基金资助(60005004)和安徽省自然科学基金资助(01042302)。王 进 硕士生,主要研究方向为网络信息处理。陈恩红 博士,副教授,主要研究方向为知识发现,网络信息处理等。林 乐 硕士生,主要研究方向为网络信息处理。

类表达式可以是一个类名、槽限制或者是使用 AND、OR 或 NOT 联接的布尔表达式。槽限制是一个或以上应用到一个槽的限制的列表。每个槽都是一个二元关系,而槽限制实际上是一个类定义,包括槽名、槽值、值类型、最大基数、最少基数和基数六个部分。槽值相当于谓词逻辑中的存在量词,对每个类的实例而言,存在至少一个槽的值满足范围约束;值类型相当于谓词逻辑中的全称量词,对类的每个实例而言,槽的每个值必须满足范围约束;基数是一个跟随了类表达式的非负整数,是槽限制所定义类的实例能通过槽被关联到的类表达式的实例数;最大和最小基数依此类推。

本体的槽定义包括七个部分,分别是定义的槽名、文档、槽继承声明、域、范围、逆槽和属性。槽名给出了槽定义的名称;文档是一些描述该类的文档;槽继承声明指出了该槽所继承的父槽表达式列表;域和范围分别定义了槽的二元关系中前后元的取值范围;逆槽定义了该槽所表示的二元关系的逆;属性的取值可以是传递的和对称的。

通过以上定义,我们就可以得到一个推荐的规范本体描述。

2.2 描述逻辑推理

描述逻辑是在知识表示研究中发展起来的,用概念(类似于类或者框架)的术语和规则(类似于框架系统中的槽)来刻画知识。在一个描述逻辑中,任何表达式的含义能被形式化的

方式刻画,从而支持了使用概念描述的推理和自动分类。我们的模型中继承了描述逻辑的形式语义和高效推理支持,使用 FaCT 系统来提供本体设计、集成和检查的推理支持。

FaCT 是一个原型描述逻辑的知识描述系统,使用了一个优化的场景包含算法来为概念描述语言提供完全推理。实际上, FaCT 主要是作为一个概念化的规划设计和本体工程工具,因此仅仅提供了基于推理服务的概念;其采用的术语逻辑 TBox 包含了公理的形式,接口的设计和 KRIS (Knowledge Representation Interactive System) 系统广泛兼容。我们继承了描述逻辑的推理原理,并针对特定的本体领域特点设计相应的推理规则 and 方式,从而在推理过程中能够高效和快捷地完成任

3. 模型的框架

在上述的理论基础之上,我们提出了一种网络环境中本体的演化和维护模型。如图1所示,该模型包括六个主要功能模块:领域本体模型管理机制、索引机制、数据抽取机制、本体演化机制、语义推理机制以及演化传播机制等。这六个功能模块结合数据库部分、知识源以及领域本体就构成了我们这个演化和维护模型的雏形。下面我们将分别对这些模块进行介绍。

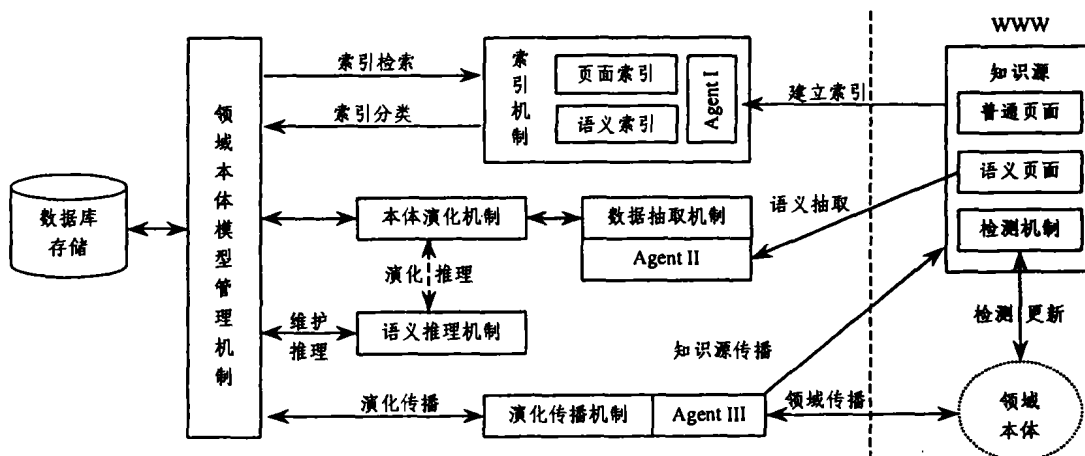


图1 本体演化和维护模型结构

3.1 领域本体模型管理和数据库存储

领域本体模型管理模块是本体演化和维护模型中的控制核心,能有效地协调其它几个模块进行协同工作,实现对知识源的索引、领域本体演化以及对演化结果进行传播等功能。同时,此模块还提供了与领域本体的语义推理机制之间的接口,能够对本体模型的语义一致性进行检查和维护。

在模型中,数据库的作用是对领域本体以及知识源的索引等进行存储以便于管理。针对本体中的槽定义、类定义和槽限制,我们分别设计了相应的表格与之一一对应。一个例子如下所示:

我们数据库的类定义表中 type、Cname、documentation、Issub 和 Sconstraint 域分别对应定义1中的类型、类名、文档、类继承声明和槽限制;其中,如果一个类的 Issub 域值为 false 的话,表示该类不是其他类的子类,否则子类关系定义表中将会对其进行说明;同样,如果 Sconstraint 域的值不为空的话,表示在槽限制表中的 Scname 域同名的槽限制所在行即是对该类的槽限制的说明。

数据库中的槽限制表中的 slotname、hasvalue、valuetype、maxcardinality、mincardinality 以及 cardinality 分别对应定义3中的类定义,包括槽名、槽值、值类型、最大基数、最少基数和基数;而 hvIsExp 表示对 hasvalue 后的内容进行判断,如果是表达式的话,则为 Yes, 否则为 No; 同理, vtIsExp 也是对 valuetype 后的内容进行说明。针对 maxcardinality、mincardinality 以及 cardinality,也有相应的表格设计。

一个槽定义包含 name、documentation、subslot-of、domain、range、inverse 和 properties,基本设计思想同前。

同时我们引入了表达式表,一元表达式如“Not”,二元表达式有“And”、“Or”和“,”。Not 表示“非”,And 表示“交”,Or 表示“并”,逗号表示使用逗号连接的多个名字列表。例如上面实例中还需要如下表进行描述。

按照这种方法,我们能够使用关系数据库来完整地对本体进行描述和存储,同时,还能很容易地生成一个关于本体中的概念和关系的多叉树,并进行概念节点的插入、删除、合并

和更新等操作,目前这部分的工作也在进一步的研究中。

一个 Ontology 的例子片段

```

slot-def eats
  inverse is-eaten-by
slot-def has-part
  inverse is-part-of
  properties transitive
class-def animal
class-def plant
  subclass-of NOT animal
class-def tree
  subclass-of plant
class-def branch
  slot-constraint is-part-of
    has-value tree
class-def defined herbivore
  subclass-of animal
  slot-constraint eats
    value-type
    plant OR
  slot-constraint is-part-of
    has-value plant
  
```

槽定义表:

Sname	Inverse	properties
eats	is-eaten-by	
has-part	is-part-of	transitive

类定义表:

Cname	type	Issub	Sconstraint
animal	primitive	False	
plant	primitive	True	
tree	primitive	True	
branch	primitive	False	Sc_branch
herbivore	defined	True	Sc_herbivore

子类关系定义表:

Cname	IsExp	Superclass
plant	Yes	Not_A
tree	No	plant
herbivore	No	animal

槽限制表:

Scname	slotname	hasvalue	hvIsExp	valuetype	vtIsExp
Sc_branch	is-part-of	tree	No		
Sc_herbivore	eats			Or_A	Yes
SC_A	is-part-of	plant	No		

图2 本体的一个实例片段在数据库中的实现

Not 表:

Expname	IsExp	Operand1
Not_A	No	Animal

Or 表:

Expname	Operand1	Op1IsExp	Operand2	Op2IsExp
Or_A	plant	No	Sc_A	Yes

图3 Not 表和 Or 表

3.2 索引和数据抽取

索引机制和数据抽取机制在模型中的任务都是获得相应操作的数据来源,通过 agent I 和 agent II 来处理知识源中的信息。它们的区别是,索引机制需要针对知识源中的普通页面和语义页面分别建立相应的页面索引和语义索引,得到的信息进行分类后存入数据库,并作为索引检索的依据;而数据抽取机制则是仅仅针对语义页面抽取其中蕴含的元数据,产生的信息作为本体模型演化的依据。

索引机制在对知识源建立好页面索引和语义索引之后,与本体模型管理模块之间的操作分为两种:索引检索和索引分类。索引检索是本体模型管理模块向索引机制发出的检索请求,按照普通页面或者语义页面进行检索目标信息;索引分类则是有本体模型管理模块对索引机制获得的索引结果进行分类的操作,当然是建立在现有领域本体模型的层次结构关系基础之上。本体模型管理机制将会对这两种操作的最终结果进行保存,并及时进行动态更新,以与知识源的当前状态保持一致。

在现有的网络环境中,网络信息主要由半结构化数据组成,因此,基于半结构化数据模式发现的元数据挖掘方法将是我们数据抽取中的重要研究对象。数据抽取机制将使用带标记有序树模型描述 Web 页面的信息的方案,生成半结构化数据文档,并根据已有的本体模型及相应的推理机制对元数

据集进行分层和归类,提供本体模型演化的数据来源。

3.3 本体模型推理、演化和传播

推理机制在整个模型中的地位举足轻重,是模型的语义核心,为索引分类、数据的分层、本体模型演化和一致性维护等提供了强大的推理支持。该模块采用了 FaCT 的核心思想,并根据特定领域设计了相应的公理体系和推理方式以提供该领域内的专门推理服务。

演化和传播机制是此模型语义维护作用的具体体现。针对领域中知识结构和表示的动态变化现象,本体模型也应该能及时作出相应的反应,演化机制就承担了这个任务。本体模型的几个基本演化过程包括对本体模型中的概念、关系、公理、域等添加、删除和修改等,在语义推理的基础上,演化机制将保证在上述的所有过程中概念和关系逻辑清晰、合理以及语义一致。同样,传播机制是为了协调领域本体、网络环境下各知识源、实例以及具体应用等与本体演化结果之间的一致性,agent III 能及时将本体演化的结果在网络环境通过传播反映出来。此外,知识源与领域本体之间也建立了检测更新的交互,间接地推动了演化结果的传播行为。

总结 本文针对领域知识的动态变化导致本体模型不能正确反映当前的现状的问题提出了一种基于网络环境中的本体演化和维护模型,能够及时地根据不断变化的领域知识进行本体模型的演化和调整以适应知识表示的需要,并对模型中领域模型管理机制、索引机制、数据抽取机制、本体演化机制、语义推理机制以及演化传播机制等六个模块进行了分析,对它们的功能作了简要的描述。在今后的工作中,我们将进一步如何设计有效的本体演化算法来实现领域本体的自学习功能,以及针对特定领域和特定用户群设计和建立个性化的本体模型,优化和整合动态知识结构的组织策略,提供全新的网络信息共享和管理机制。

(下转第5页)

object O 到 FOL 的 (PropertyValue P S O)。这里的 PropertyValue 表示一个三元关系。

(3)为 RDF, RDF Schema 和 DAML+OIL 的 property, class 和 constant 增加相应的公理(Axiom)。

DAML+OIL 的紧凑语法是不同于 (property, subject, object) 三元组形式。因此第一步工作就是将这样的紧凑语法转换成 RDF 的 statement。在这个步骤中只需要将 RDF statement 中的每个元素用 URI 或 literal 标记, 或者由转换器为其生成一个唯一的 label。

在第二步中, 将 DAML+OIL 的声明 (property P of resource R has value V) 用 KIF^[20] (PropertyValue P R V) 来表示。因为 type 属性是 RDF 的核心, 可以在 KIF 中另外定义一个二元关系 Type 来支持这个转换。

($\langle = \rangle$) (Type ? r ? v) (PropertyValue type ? r ? v))

它的含义是 RDF 声明 “property P of resource R has value V” 等价于 “Property ‘type’ of resource R has value V”。在 KIF 中 ($\langle = \rangle$) 的含义是 “当且仅当”。

第三步中为 RDFS 和 DAML+OIL 中的 class 和 property 词汇增加相应的公理描述, 参见文[21]。

将 DAML+OIL 转换成 FOL 后, 就是可以利用 FOL 成熟的推理机制来进行逻辑推理, 由此语义 Web 就具备了逻辑推理能力。

结束语 Tim Berners-Lee 在 2001 年描绘出了一幅语义 Web 的蓝图。在未来的语义 Web 中, 人们可以利用 Agent 来完成目标信息的自动查找, 计划的安排, 与其他 Agent 进行交流等工作。

语义 Web 中使用的 Ontology 语言也在不断地补充和发展。继 DAML+OIL 语言之后, 2002 年, W3C 公布了 Web Ontology Language (OWL) 工作草案, 该草案提出了 OWL 语言, 目前该草案仍然处于讨论和修改之中, 相信不久就会成为 W3C 新的标准。

参考文献

- 1 Berners-Lee Tim, Fischetti M. Weaveing the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor. San Francisco: Harper. 1999
- 2 Bray T, Paoli J, Sperberg C. Extensible Markup Language (XML) 1.0. W3C Recommendation. Available at: <http://www.w3.org/TR/REC-xml>. 1998

- 3 Brickley D, Guha R V. Resource Description Framework (RDF) Schema Specification. W3C Candidate Recommendation. Available at: <http://www.w3.org/TR/PR-rdf-schema>. 2000
- 4 Horrocks I, Fensel D, Harmelen F, Decker S, Erdmann M, Klein M. OIL in a Nutshell. In: 12th Intl. Conf. in Knowledge Engineering and Knowledge Management, Lecture Notes in Artificial Intelligence, 1-16. Berlin, Germany. Springer-Verlag. Available at: <http://www.cs.vu.nl/~ontoknow/oil/down/oilnutshell.pdf>. 2000
- 5 Horrocks I, van Harmelen F. Reference description of the DAML+OIL ontology markup language. Draft report, 20001. Available at: <http://www.daml.org/2000/12/reference.html>. 2001
- 6 Kent R. Conceptual Knowledge Markup Language (Version 0.2). <http://sern.ucalgary.ca/KSI/KAW/KAW99/papers/Kent/CKML.pdf>. 1998
- 7 Luk S, Heflin J. SHOE 1.01. Proposed Specification. SHOE Project. Available at: <http://www.cs.umd.edu/projects/plus/SHOE/spec1.01.htm>. 2000
- 8 SHOE. <http://www.cs.umd.edu/projects/plus/SHOE/KnowledgeAnnotator.html>. 2002
- 9 Web Ontology Language (OWL) Reference Version 1.0. W3C Working Draft 12 November 2002. Available at: <http://www.w3.org/TR/2002/WD-owl-ref-20021112/>. 2002
- 10 Connolly D, van Harmelen F, Horrocks I, et al. DAML+OIL (March 2001) Reference Description. W3C Note 18 December 2001. Available at: <http://www.w3.org/TR/daml+oil-reference>. 2002
- 11 Fensel D. The Semantic Web and its languages [J]. IEEE Intelligent Systems, 2000(11/12): 67~73
- 12 Berners-Lee T, Connolly D, Swick R. Web Architecture: Description and Exchanging Data. Available at: <http://www.w3.org/1999/04/WebData.html>. s
- 13 Berners-Lee T. Semantic Web-xml2000-Slide “Architecture”. Available at: <http://www.w3.org/2000/?Talks-xml2k-tbl/Overview.html>
- 14 Decker S, Mitra P, Melnik S. Framework for the Semantic Web: A RDF Tutorial [J]. IEEE Internet Computing, 2000(11/12): 68~73
- 15 Bechhofer S, et al. An informal description of stanford OIL and instance OIL. Available at: <http://www.ontoknowledge3.org/oil/>
- 16 Hendler J. Agents and the Semantic Web [J]. IEEE Intelligent System, 2001(3/4)
- 17 Berners-Lee T, Fielding R, Masinter L. Uniform Resource Identifiers (URI): Generic Syntax. RFC 2396, Aug. 1998
- 18 Berners-Lee T. Web Architecture from 50,000 feet. Available at: <http://www.w3.org/DesignIssues/Architecture.html>
- 19 Berners-Lee T. Semantic Web Road map. Available at: <http://www.w3.org/DesignIssues/Semantic.html>
- 20 Knowledge Interchange Format (KIF) Available at: <http://logic.stanford.edu/kif/kif.html>
- 21 McGuinness D L, Fikes R, Hendler J, et al. DAML+OIL: An ontology language for the semantic web. In: Proc. of the 18th Nat. Conf. on Artificial Intelligence (AAAI 2002), 2002

(上接第 128 页)

参考文献

- 1 Lassila O, Swick R R. Resource description framework (RDF) Model and syntax specification. Recommendation, W3C, February 1999. <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222>
- 2 W3C. RDF Vocabulary Description Language 1.0: RDF Schema. W3C Candidate Recommendations are available at <http://www.w3.org/TR>, 30 April 2002. Edited by Dan Brickley and R. V. Guha
- 3 Horrocks I, Fensel D, Goble C, et al. The ontology inference layer oil: [Technical report]. Free University of Amsterdam, 2000. <http://www.ontoknowledge.org/oil/>
- 4 Decker S, Fensel D, van Harmelen F, et al. Knowledge representation on the web. In: Proc. of the 2000 Intl. Workshop on Description Logics (DL2000), Aachen, Germany, 2000
- 5 Horrocks I, Sattler U. Ontology reasoning in the SHOQ (D) description logic. In: B. Nebel, ed. Proc. of the Seventeenth Intl.

Joint Conf. on Artificial Intelligence (IJCAI-01), Morgan Kaufmann, Los Altos, 2001. 199~204

- 6 Horrocks I. A denotational semantics for Standard OIL and Instance OIL, 2000. <http://www.ontoknowledge.org/oil/down/semantics.pdf>
- 7 Horrocks I, Sattler U, Tobies S. Reasoning with individuals for the description logic shiq. In: Proc. of CADE-17, number 1831 in LNCS, Springer-Verlag, 2000. 482~496
- 8 Franklin S, Graesser A. Is it an Agent, or just a Program: A Taxonomy for Autonomous Agents. In: Proc. of the Third Intl. Workshop on Agent Theories, Architectures, and Languages (ATAL-96), Budapest, Hungary, 1996. 193~206
- 9 Stojanovic L, Stojanovic N, Handschuh S. Evolution of the Metadata in the Ontology-based Knowledge Management Systems. In: Mirjam Minor, Steffen Staab, eds. Proc. 1st German Workshop on Experience Management, March, 2002
- 10 Doran J E, Franklin S, Jennings N R, Norman T J. On cooperation in multi-agent systems. The Knowledge Engineering Review, 1997, 12(3). Available at: <http://www.elec.qmw.ac.uk/dai/pubs/fomas.html>