

# 语义 Web 中的核心层技术研究

王 利 李吉桂

(华南师范大学计算机系 广州 510631)

**摘 要** 文中在简要介绍语义 Web 概念的前提下,针对语义 Web 模型中核心层的相关技术如 XML、RDF 特别是本体技术进行研究,并分析了各种技术之间的相互关系。所进行的研究对语义 Web 的构建和实现具有一定的促进作用。

**关键词** 语义 Web, XML, RDF, 本体

## 1 引言

语义 Web 是由 Web 之父 Tim Berners - Lee 提出的新一代 Web,其目标是使得 Web 上的信息具有计算机可以理解的语义,从而可以通过智能软件代理(Agent)对 WWW 上异构和分布的信息进行有效的检索和访问。

语义 Web 的层次如图 1 所示<sup>[1]</sup>:

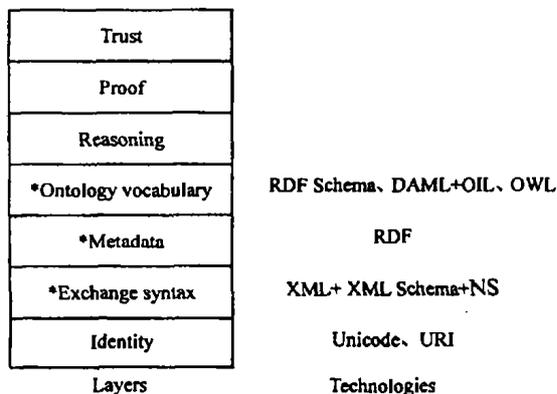


图 1 语义 Web 的层次

其中:Unicode、URI 是整个语义 Web 的基础,用于资源的编码和标识;交换语法、元数据以及本体词汇层是语义 Web 中的核心层,用于表示 Web 信息的语义;推理层主要定义逻辑描述原语,为更高层的逻辑推理提供语义上的支持;最上面两层则是在下面各层提供的功能基础上进行逻辑推理、证明等操作。

本文对语义 Web 中核心层的有关技术特别是本体技术以及各种技术之间的相互关系进行分析。

## 2 XML + XML Schema + NS

XML 作为 Web 的知识表示语言,其核心在于以一种标准化的方式来建立数据表示的结构,而将具体标记的定义留给了用户,从而克服了 HTML 语言中必须使用指定标记带来的不便。因为使用同样的标记表达不同的语义会混淆不同的领域知识,所以 XML 提供了命名空间(NameSpace)。同时为了

便于数据交换 XML 提供了一种约束机制,用来限定 XML 文档所用到的标记和这些标记之间的结构关系。最早 W3C 采用文档类型定义(Document Type Definition, DTD),然而 DTD 本身采用与 XML 完全不同的语法,不支持命名空间,而且对数据类型的支持极弱。因此 W3C 开发了 XML 模式(XML Schema, XMLS),XML 模式采用了更为规范的方法定义 XML 文档结构,同时,提供了比 DTD 更丰富的数据类型。

XML 模式为 XML 文档提供了一定的语义描述能力,然而 XML 模式的语义仍然是隐含的。XML 模式并不能对其所含有的语义进行任何解释。为了实现 XML 文档的解释或者转换,用户必须将这些语义编码到工具中,而将数据的最终解释留给了特定的应用。

因此 XML 在形式上统一了信息交换的语法,但在语义操作性方面仍然存在不足。

## 3 RDF

RDF 是为弥补 XML 无法为信息提供语义而提出的。RDF 是一种描述和交换元数据的框架,它提供了应用系统之间交换机器可理解信息的支持,从而提高了系统的互操作性,并且在表达语义方面,RDF 提供了一种统一的数据模型。在这种模型中,任何可以通过 URI 表示的对象都是 RDF 资源。资源之间可相互联系或通过属性 - 值对联系,资源 - 属性 - 值三元组构成一个声明(Statements)。RDF 的数据模型实质上是一种二元关系的表达,由于任何复杂的关系都可以分解为多个简单的二元关系,因此 RDF 的数据模型可以作为其他任何复杂关系模型的基础模型,从而为 Web 上的知识表示、知识推理和知识校验提供了底层的标准。

RDF 和 XML 是互为补充的。首先,RDF 希望以一种标准化,互操作的方式来规范 XML 的语义。RDF 可以帮助解析器在阅读 XML 的同时,获得 XML 所要表达的主题和对象,并可以根据它们的关

系进行推理,从而做出基于语义的判断。XML 的使用可以提高 Web 数据基于关键词检索的精度,而 RDF 与 XML 的结合则可以将 Web 数据基于关键词的检索更容易地推进到基于对象的检索。

其次,由于 RDF 是以一种建模的方式来描述数据语义的,这使得 RDF 可以不受具体语法表示的限制。虽然 RDF 数据模型既可以用 Notation3<sup>[2]</sup>来表示,也可以用 XML 来表示。但是,由于 XML 已经成为被广泛支持的 Web 数据表示标准,便于应用的读取,因此将 RDF 序列化为 XML 表示可以使 RDF 获得更好的应用可处理特性,并使得 RDF 数据可以像 XML 数据一样的容易使用、传输和存储。

因此,RDF 是定制 XML 的良伴,而不只是对某个特定类型数据的规范表示<sup>[3]</sup>,XML 和 RDF 的结合,不仅可以实现数据基于语义的描述,也充分发挥了 XML 与 RDF 的各自优点,便于 Web 数据的检索和相关知识的发现。

RDF 数据模型提供了领域中立的元数据描述方法,而 RDF Schema 提供了定义不同领域核心词汇及其关系的机制。

#### 4 RDF Schema

RDF 数据模型就命名属性和值而言,定义了描述资源间相互关系的一种简单的模型。RDF 属性可被认为是资源的属性,RDF 同样代表了资源间的关系。因此 RDF 数据模型和实体-关系图类似。但是 RDF 数据模型没有提供机制来说明这些属性,也没有提供任何机制来定义这些属性和其它资源间的关系。RDF 模式进一步来完善此框架。

RDF 模式(RDF Schema, RDFS)提供了一个定义在 RDF 之上的抽象词汇集,如: rdfs: Resource、rdfs: Class 和 rdf: Property 类, rdf: type、rdfs: SubClassOf 和 rdfs: SubPropertyOf 属性关系,以及 rdfs: domain 和 rdf: range 的属性约束关系。该词汇集构成了一种基本类型系统。

RDFS 仍是一种简单的知识表示语言,它不能提供丰富的语义。例如 RDF 没有提供局部的 range 和 domain 约束,没有 transitive, inverse or symmetrical 属性,没有 existence/cardinality 约束,难以提供推理支持等等<sup>[4]</sup>。

#### 5 OWL

本体(Ontology)最早是一个哲学的范畴,后来随着人工智能的发展,被人工智能界给予了新的定义。在计算机领域,本体被定义为“共享概念模型的明确的形式化规范说明”(1998/Studer)。本体通过对概念的严格定义和概念之间的关系来确定概念精确含义,表示共同认可的、可共享的知识。在语义

Web 中,本体具有非常重要的地位,是解决语义层次上 Web 信息共享和交换的基础。

如前面所述,RDFS 在 RDF 的基础上提供了更多建模原语用于元数据定义,这些原语对进一步构造本体有一定作用。事实上,RDFS 所构造的元数据已经是一种轻型的本体知识,不过 RDFS 的功能仍然很有限。为了构造更加完备的本体信息以支持自动推理,需要更加完备的本体描述语言。

OWL 是以 DAML + OIL 的研究成果为基础所发展而得的标准,其目的是希望提供一个标准的语言,用以描述 Web 中所用到的实体类别(Class)以及各类别之间的关联性,借以加强 Web 环境中信息的语意内涵。

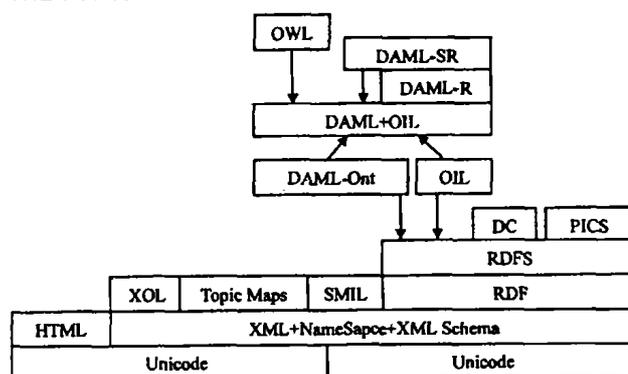


图 2 本体语言“栈”

OWL 是 W3C 本体语言“栈”中的一部分,如图 2 所示。OWL 建立在 RDF、RDFS 以及描述逻辑的基础(Description Logic)之上,通过增加更多建模原语来描述属性、类以及它们之间的关系。OWL 采用面向对象的方式来描述领域知识,即通过类和属性来描述对象,并通过公理(Axioms)来描述这些类和属性的特征和关系。

OWL 包括三种子语言<sup>[5]</sup>:

\* OWL Lite:提供分类层次和简单的约束。

\* OWL DL:在满足计算完全性(computational completeness,所有的结论都保证是可计算的)和可判定性(decidability,所有计算可在有限时间内完成)前提下,提供最大程度的知识描述能力,包括 OWL 语言的所有 constructs,只能在特定的约束下使用。OWL DL 与描述逻辑相一致。

\* OWL Full:能够提供最大程度的知识描述能力和 RDF 的语法自由,没有计算上的保证。OWL Full 允许本体扩展现有的 RDF 或 OWL 词汇。任何推理软件都不太可能完全支持 OWL FULL。

在使用 OWL 进行本体的开发时,具体使用哪一种子语言取决于对语言描述能力以及对 RDF 模式灵活的元模型机制(如定义类型的类型或为类型赋予属性)的需求。OWL Full 可以看作是 RDF 的扩展,OWL Lite 和 OWL DL 则可以看作是一个约束化的 RDF 的扩展,目前还没有完全的 OWL Full

实现。

OWL 语言的表述能力体现在对类和属性的描

述上,即通过一系列的 constructors 来构造类和特性。如表 1、表 2 所示。

表 1 OWL Lite 语言

constructors	terms
RDF Schema Features	略
(In)Equality	equivalentClass、equivalentProperty sameIndividualAs、differentFrom、allDifferent
Property Characteristics	inverseOf、TransitiveProperty、SymmetricProperty FunctionalProperty、InverseFunctionalProperty
Property Type Restrictions	allValuesFrom、someValuesFrom
Restricted Cardinality	minCardinality、maxCardinality、cardinality
Header Information	Ontology、Imports
Versioning	VersionInfo、priorVersion、backwardCompatibleWith...
Class Intersection	intersectionOf
Annotation Properties	略
Datatypes	xsd datatypes

表 2 OWL DL 和 OWL Full 语言

constructors	terms
Class Axioms	one of、disjointWith...
Boolean Combinations of Class Expressions	unionOf、intersectionOf、complementOf
Arbitrary Cardinality	minCardinality、maxCardinality、cardinality
Filler Information	hasValue

从上面的表中可以看出,OWL 在 RDFS 构建的基本类型系统的基础上,进行了如下几方面的扩展:

### 1. 头信息和版本

OWL 支持本体包含以及版本管理。包括了用于指定导入的本体 (Imports)、本体版本信息 (VersionInfo) 和前版本信息 (priorVersion) 等一系列信息的方法。

### 2. 属性

OWL 中有两类属性:数据类型属性 (datatype property) 和对象属性。前者联系对象和数据类型值,后者联系两个对象。

OWL 使用 XML Schema 提供的数据类型 (xsd datatypes), 这些数据类型能够通过 URI 被识别。之所以不在 OWL 中定义数据类型特性,而引用 XML Schema 中定义的数据类型,并且将这种数据类型和 OWL 中定义的对象相区分,是因为:

1) 已知的数据类型已经足够丰富<sup>[6]</sup>, 因此没有必要利用本体语言定义新的数据类型;

2) XML Schema 中定义了大量的数据类型,如果使用本体语言重新定义这些数据类型,将加大语言的复杂性,这有悖于语言定义的简单性原则;

3) 使用引入 XML Schema 中数据类型的方法增强了本体语言的扩展性,使得本体语言能够相对容易地应用在已经采用了 XML Schema 定义数据类型的应用。

### 3. 属性特征

OWL 扩展了属性之间的关系, Symmetric Property、Transitive Property 和 inverseOf 定义了属性之间的传递、对称和相反关系; Functional Property 和 Inverse Functional Property 则提供了更加丰富的属性描述。Functional Property 声明一个属性只有一个值, Inverse Functional Property 声明一个属性的相反属性是 Functional 的。这些机制提供了有效的推理支持。

### 4. 属性约束

除了全局约束, OWL 还提供了以下两个局部约束: allValuesFrom 和 someValuesFrom 约束将一个属性的取值和一个类相关。另外 minCardinality、maxCardinality、cardinality 约束可用来确认一个类实例最多(少)具有多少某属性的实例(集的势)。  
[注: OWL Lite 在 Cardinality 上的局部约束只允许

(下转第 90 页)

- 9 王桂玲,许骏,等.G-CSCLE:基于 OGSA 的协作学习环境的实现. 见:2003 中国计算机大会论文集,清华大学出版社,2003.1376~1380
- 10 Yang Wei, Yuan Rong. The Research and Design of Intelligent E-learning System Computer Science and technology in New Century. International Academic Publishers World Publishing Corporation. Oct.2001.1245~1247

- 11 杨威.一种用于网上学习的智能答疑模型. 计算机工程, 2003(12):173~175
- 12 Canntaro M, Talia D, Trunfio P. KNOWLEDGE GRID: High Performance Knowledge Discovery Services on the Grid. proc GRID 2001. LNCS, Springer-Verlag, 2001. 38~50

(上接第 85 页)

Cardinality 的取值为 0 和 1, 不同于其它两类 OWL 允许任意数目的 Cardinality。hasValue[OWL DL] 约束允许我们定义基于特定属性值存在的类: 一个实例要成为一个类的成员, 它在这个属性上的取值必须满足 hasValue 的规定。

### 5. 本体映射的支持

equivalentClass 和 equivalentProperty 定义了类和属性之间的等价关系; sameIndividualAs、differentFrom、allDifferent 定义了实例之间的相同、不同以及两两不同关系。通过它们可在 Ontology 上定义映射关系, 以重用已有的 Ontology 的类和属性。

### 6. 定义复杂的类

OWL 通过定义集合操作符 Intersection[OWL DL]、Union[OWL DL]、Complement[OWL DL]、枚举操作符 one of[OWL DL] 以及类之间的不相交关系 disjointWith[OWL DL] 给出了定义类表达式的基本方法, 从而能够通过嵌套定义给出一个复杂的类。

此外 IntersectionOf 允许在类和约束之间存在交集。

**总结与展望** 语义 Web 是当前的一个研究热点。而要实现语义 Web, 其中重要的一步就是为 Web 资源赋予机器可读的语义信息。因此笔者根据 W3C 提出的一系列标准对语义 Web 模型中核心层的相关技术进行了研究, 包括 XML、RDF 和 OWL 各自的性质和作用, 以及它们之间的相互联系。

目前, 语义 Web 仍处在构想阶段, 存在着许多问题有待于进一步的研究和解决。如基于 RDF 的数据挖掘和本体管理(如合并, 映射, 进化), 以及如

何进一步在此基础上实现更加智能化的搜索引擎、电子商务和知识管理, 如何将语义 Web 与 Web 服务相结合, 提供智能化的 Web 服务等等。

### 参考文献

- 1 Editorial G. The Semantic Web an evolution for a revolution. Computer Networks, 2003, 42:551~556
- 2 Sean B P. The Semantic Web: An Introduction [EB/OL]. <http://infomesh.net/2001/swintro>, 2001-09
- 3 Ogbuji U. Generate RDF using XSLT. <http://www-106.ibm.com/developerworks/library/x-think4/>, 2001
- 4 Bechhofer S, Horrocks I, Peter F. Patel-Schneider. Tutorial on OWL
- 5 McGuinness D L, van Harmelen F. OWL Web Ontology Language Overview. <http://www.w3.org/TR/2004/REC-owl-features-20040210/> 10 February 2004. OWL Web Ontology Language Overview. <http://www.w3.org/TR/2004/REC-owl-features-20040210/>
- 6 Horrocks I, Patel-Schneider P F, van Harmelen F. Reviewing the Design of DAML + OIL: An Ontology Language for the Semantic Web [A]. In: Proc. of the Eighteenth National Conf. on Artificial Intelligence and Fourteenth Conf. on Innovative Application of Artificial Intelligence [C]. Edmonton, Alberta, Canada: AAAI Press, 2002. 792~797
- 7 周竞涛, 王明微. XML+RDF--实现 Web 数据基于语义的描述. <http://www-900.ibm.com/developerWorks/cn/xml/x-xmlrdf/index.shtml>
- 8 OWL Web Ontology Language Guide <http://www.w3.org/TR/2004/REC-owl-guide-20040210/> 10 February 2004, Michael K. Smith, Chris Welty, Deborah L. McGuinness
- 9 Ding Y, Fensel D, Klein M, Omelayenko B. The semantic web yet another hip