

Semantic Web

语义 Web 及层次结构

and its Layer Architecture

摘要: 本文描述了下一代万维网——语义 Web 及其基本内容, 文中介绍了语义 Web 的定义及其层次结构, 描述了支持语义 Web 各层的基本技术如 RDF (资源描述框架) /RDF 模式语言 OIE(本体推理层), 逻辑层及有关技术等, 并描述了语义 Web 的发展前景。

关键词: 语义 Web 层次结构 模式 本体 逻辑

1 语义 Web 结构

Berners-Lee 给出如下定义: 语义 Web 是一个网, 它包含了文档, 或文档的一些部分, 描述了事物间的明显关系, 且包含语义信息, 以利于我们的机器自动处理。

在 XML 2000 大会的重要发言中, Berners-Lee 专门解释了“语义”和“Web”两个词的含义。在“Web”背后是可导航空间的思想, 用 URI (统一资源标识符) 映射到资源上。语义 Web 所指的“语义”是“机器可处理”, 而不是自然语言语义和人的推理。对于数据, 语义表达了机器能对这些数据做些什么, 若你给出一些数据到一台机器, 它用这些数据做出正确事情, 就称它通过了“语义测试”, Berners-Lee 描述了他对语义 Web 结构的设想, 他预见这个层次结构将在今后十年

内得到发展。其结构图如图 1 所示。

其中, Unicode 是一个新的字符编码系统标准, 支持世界上所有主要语言文本的混合; URI 是 Internet 资源的一种识别方法, XML 于 98 年由 W3C 设计出来, 主要是用于使信息自描述, 准确告诉机器信息的

含义, XML 包含一些规则, 任何人可按规则创建一个标记语言, 规则保证一个称为分析器的简洁程序可以处理这些新语言, NS 是 XML 名称空间, 由 URI 索引确定, 在 XML 文档中用作元素类型和属性名, XML 模式 (schema) 是描述 XML 文档内容的模型, XML 模式内有两类基本约束: 内容约束 (它决定在哪里及什么时候可以使用元素) 和数据类型约束 (它控制什么类型数据可以出现在元素中), XML 依赖于 Unicode, 因此它能在不同计算机系统间交换信息, 也超越了国家和文化的边界。

RDF 可提供一种处理元数据的环境, 元数据即描述数据的数据, 这些描述包含需要计算机理解的数据, RDF 增加语义到文档中而不用假定文档的结构, 其语法可用 XML 定义, 万维网上的数据是机器可读取的, 但不一定是机器可理解的, 如果网上的数据遵循一种标准的描述或标记方法, 那么计算机对于它们处理的资源一定会理解得更多,

所以 RDF 标准为语义 Web 奠定了基础, RDF 模式引入一个面向对象、可扩展的类型 (type) 系统到 RDF 中, 提供方法定义合适的定义域和值域, 以及类和子类层次。

Ontology 译为本体, 是某领域内概念的显式说明和对应的词汇用于描述该领域, 一个本体可以描述事情的类型之间的关系, 如“这是一个过渡的性质”, 本体词汇提供更多元信息, 互操作性和互转换性, 信息的无缝交换已成为 Web 成功开发的关键问题, 本体提供了方法捕捉人和机器使用的术语的共享理解性, 帮助信息交换。

逻辑层是利用万维网各处的断言导出新知识的地方, 问题是各种演绎系统不可互操作推理能力不同, 任何规则系统都可以输出到这层。

如果不设计一个跨系统的推理系统, 而是建议一种普遍性语言来表达证明, 就实现了证明层, 数字签名跨越了多层, 虽然公共密钥密码术已存在多时, 但还没有真正起

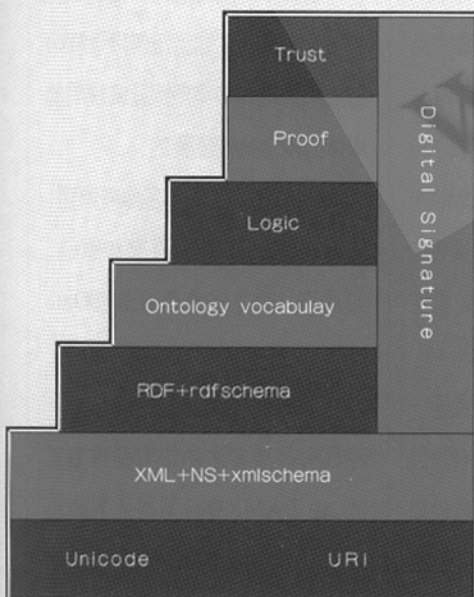


图 1 语义 Web 结构

飞,如果它真正实用,加上语义Web 各层支持,使一个团体在一定范围内可信任,就实现了信任层。重要的应用就可以进入语义Web,如电子商务等。

2 RDF 模式层

RDF是W3C新建的标准,通常用于表达元数据,描述其他Web资源如XML文件。它提供一个数据模型,利用跨越语义差别支持快速数据集成。它又像最小公分母,建立Web应用程序之间的互操作性。使用标准方法处理这种元数据,人们就能以对描述语言预期的理解来设计应用程序、句法和涉及的传输,使得应用程序能够交换信息而不必操心互操作性问题。

RDF数据模型包括三个基本组成部分:资源(能通过URI引用的任何事物)、属性(被描述资源的特性)和语句(包括引用资源的指针以及该资源属性和属性值的表达式)。RDF语句可以用直接标记图来表达,但为了机器交换,还需要适当的句法,比如可以用XML作为其句法框架。RDF也依赖于XML规范中的名称空间,以正确地把属性固定在定义它们的模式上。假如要表示如下语句:

LiWeihua是<http://www.gdut.edu.cn/index.html>的作者。

用XML句法写成的RDF语句就变成如下形式:

```
<rdf:RDF>
<rdf:Description about=http://www.gdut.edu.cn/index.html>
```

```
<Author>LiWeihua</Author>
```

```
</rdf:Description>
```

```
</rdf:RDF>
```

在语义Web中,RDF是作为普遍的语义断言,最小设计原则要求它非常一般。这样一来,任何有希望的应用就可以映射到该模型中。由于一般,它很简单,RDF基本模型没有给我们多少工具,因此需要一个模式层声明新性质的存在。模式的概念来源于数据

RDF实例数据工作。

RDF模式说明书在2000年3月成为W3C的候选推荐,这个类型系统的元结构是URI命名的术语和概念,所以RDF有效地表达和定义类和属性。类定义可从多个超类中导出,属性定义可详细叙述定义域和值域的限制。

模式说明书提供了核心类、核心属性类型和核心限制,三个核心类由RDF模式机构提供:资

本体层将RDF模式扩展为齐全的本体建模语言。

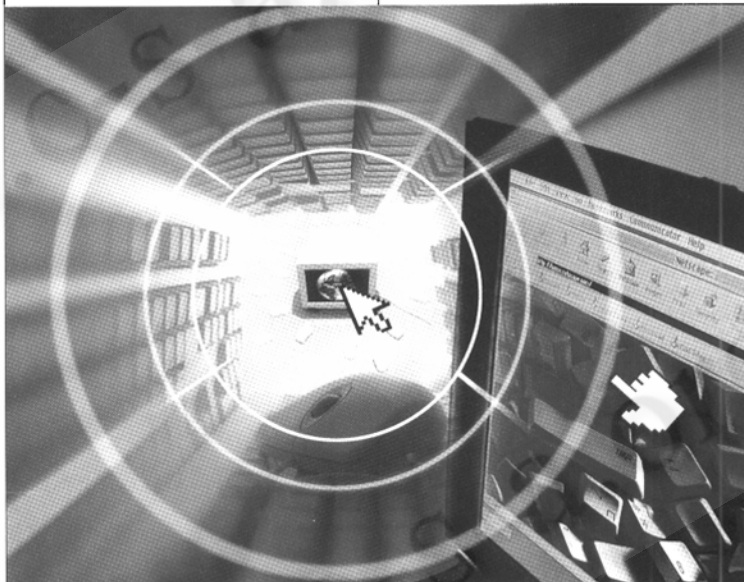
3 本体层

本体最早是哲学概念,即“存在的研究或科学”,后来Gruber定义它为“概念的详细说明”。现在,许多领域都引入本体的概念,如知识工程、知识表达、语言工程、数据库设计、信息存取、知识管理等。粗略地说,本体相当于一般化的数据库模式,但是,本体可以用于描述比普通数据库更复杂的对象的语义结构。因此,它很适合于描述异种的、分布的、半结构的信息资源,如同Web上的信息。

为了在不同的应用程序之间共享信息和知识(用于互操作),需要有一组共享的具有公共理解的术语,描述应用程序的领域,而且定义这些术语间的关系,这一组术语就是本体。

在语义Web团体内,本体是一组知识术语,包括词汇、语义互连和一些简单的推理规则与逻辑,用于某个特定的主题。任何用于Web的语义标记方案,支持显式说明的本体的概念是很重要的。

本体需要某种语言来说明,现在已有一些表达语言和系统,如基于一阶逻辑的表达语言Onto lolingua、Loom、Frame-Logic等。但是,对于Web上的应用,有一种具有标准句法的语言是很重要的。由于XML有望成为Web上数据交换的标准语言,所以也希望用XML句法交换本体,这样会



库,在那里模式用于描述表和构成存储数据的表内的字段。RDF模式引入一个面向对象、可扩展的类型系统到RDF中,提供方法定义合适的定义域和值域,以及类和子类层次,通常我们需要限制它能加入的对象的类型。

在RDF模式中,属性不是局部于一个类(如面向对象语言中的属性),而是全局的,在属性连接的各类中描述。模式文档的作用类似一个合同作用:在提交一个RDF模式的同时,应用程序开发者保证他们的应用程序可与依从RDF模式的

源(即所有对象的类),属性类型(即所有二元关系的类),类(即所有类型的类)。有两个核心属性类型:instanceOf(定义一个资源与类的一个元素间的关系)和subClassOf(定义类的两个元素间的关系)。限制有两个核心实例:range和domain,分别定义属性类型的值域和定义域。

我们可以认为RDF模式是一组本体建模原语加在RDF上[1],但RDF模式提供的语言比其他现代知识表达语言要弱,比如,没有标准方法描述基本限制,有待其上层的

简化解析程序的任务。这种需求引导了基于XML的表达语言,在XML上定义一层语言,如SHOE、本体交换语言(XOL)、本体标记语言(OML和CKML)、RDF模式语言、Riboweb等。

本体互换语言OIL(Ontology Interchange Language)是扩展RDF和RDF模式的新建议,它以RDF模式作为起点,扩展它为完整的本体建模语言。OIL也称本体推理层(Ontology Inference Layer),建议了一种层次的标准本体语言方法,能够对许多知识表达语言结构进行建模,与RDF模式兼容,并包含精确语义用于描述术语的意义。

OIL有核心OIL、标准OIL、实例OIL和重型OIL四层,核心OIL大部分与RDF模式相符合,标准OIL用于捕捉必要的建模原语主流,既提供足够的表达能力,又容易理解,因而允许语义精确定义和完全推论可行,实例OIL包括完整的数据库能力,有彻底的个体集成,重型OIL用于将来的扩展,加入其他表达和推理能力。

OIL是欧盟信息团体的推荐技术,另一个基于Web的本体建模语言候选者是美国DARPA的DAML(DARPA Agent Markup Language)技术,DAML继承了OIL的许多方面特点,两种语言的能力大致相同。这两个团体组成一个联合会,吸取了两种语言和其他一些语言的优点,形成了DAML+OIL,这是目前最先进的Web本体语言,有望成为将来Web本体标准。

4 逻辑层

语义Web的逻辑层为单调逻辑提供通用的语言,任何规则系统可以输出到该层,但一般不能从该层输入,目前没有一个标准的推理机,各种推理机有不同的推理能力,如SQL to KIF、Cycl等系统[3]。

为了处理语义Web中可得的知识,需要有一个通用的推理机,从已经定义的知识中导出新知识。建造语义Web上的推理机有两种不同的可行方法:基于一般逻辑的推理机,和专用算法(问题求解方法)。

对于第一种方法,有不同的表达语言和推理机,如基于高阶逻辑的推理机,完全一阶逻辑的推理机,描述逻辑、Datalog与逻辑编程。

高阶逻辑是所有已知逻辑中表达能力最强的,但是,它没有好的计算性,存在真命题,但不可证明,高阶逻辑有高阶句法和高阶语义两方面,如果不需要高阶语义,二阶逻辑可以转换为一阶逻辑。

用完全一阶逻辑描述公理需要完备的自动定理证明器,一阶逻辑是半可判定的,对大量数据和公理进行推理在计算上是无法容易控制的,这意味着在Web这样的环境下,这类程序不能扩大以处理大量知识。

描述逻辑允许用一组受限制的一阶公式描述术语的层次,有好的计算性,但推理服务限于分类和归类。

另一种用于语义Web的推理机的可能性是基于Horn逻辑的语言,Horn逻辑与Datalog用于演绎数据

库和逻辑编程领域,从分布在Web上不同资源中集成Horn逻辑会导致困难,各规则会互相干扰,导致不希望的结果。

对于第二种方法,问题求解方法是小的算法,用于知识库系统和知识获取领域,在专家系统内执行推理。

目前,逻辑层的工作才刚刚开始,DARPA的DAML的下一个目标是创建一种逻辑语言的早期版本:DAML逻辑,具有足够的方法表达公理和规则^[1]。我们期待DAML逻辑能实现语义Web逻辑层的作用。

5 语义Web发展前景

语义Web虽然刚起步,还要经过多年的努力才能显出威力,但它却有重要的发展前景。近期内几个主要应用领域是^[1]。

5.1 搜索引擎

当前的搜索引擎受依赖关键词匹配的严重影响,不能找出用不同术语描述的相关信息,也不能区别使用同一个词而意义不同的信息,如果搜索引擎可以搜索网页信息的语义概念,将会解决这个问题。

5.2 电子商务

当前用户只能用人工访问每个网上商店以进行比较,软件代理只能帮助用户搜索网页的简单信息(如价格),商店网页显示格式一变,软件代理就要更新。若网上商店提供机器可处理的目录,软件代理就可以代替用户访问网上商店。

5.3 知识管理

越来越多企业依靠intranet技术为其员工作为知识存储处,传统的文件管理系统没有足够的方法构造和存取网上知识,语义Web能提供机制将文件存储转变成合适的知识存储。

6 结论

下一代万维网已被称为语义Web,其信息将是机器可处理的方式,支持智能网络服务,如信息中介和搜索代理,语义Web要求互操作性标准不仅表达文档的构造形式,还要表达其语义内容,W3C的XML/XML模式和RDF/RDF模式正朝语义互操作性的标准化方向努力,还需要更多团体参与工作,语义Web又是开放的系统,允许各独立团体加入各种技术,除了W3C(World Wide Web Consortium)的标准外,还有欧盟信息团体的OIL技术和美国DARPA的DAML技术等。 ■

参 考 文 献

- 1 Fensel D. The Semantic Web and its languages. IEEE Intelligent Systems, 2000,Nov./Dec:6773.
- 2 Berners-Lee T, Connolly D and Swick R. Web Architecture: Describing and Exchanging Data. Available at <http://www.w3.org/1999/04/WebData.html>.
- 3 Berners-Lee T. Semantic Web - XML2000 - Slide rchitecture Available at <http://www.w3.org/2000/Talks/xml2k-tbl/Overview.html>.