doi:10.3969/j.issn.1000-436x.2012.z2.007

语义 Web 安全研究

陈德彦[1,2]，赵宏[1]，张霞[1,2]，赵立军[2]，平安[2]
(1. 东北大学 计算机软件国家工程研究中心，辽宁 沈阳 110179；
2. 东软集团股份有限公司 软件架构国家重点实验室，辽宁 沈阳 110179)

摘 要：针对语义 Web 技术本身以及由遍及 Internet 各个角落的语义 Web 应用构成的语义数据网可能引入新的安全、隐私和信任问题及需求，详细分析了语义 Web 所面临的安全问题及其安全现状。在此基础上，提出了语义 Web 的安全参考技术框架。最后，指出在该安全框架下语义 Web 安全急需研究的关键技术、策略规范和安全标准。
关键词：语义 Web；本体；语义 Web 安全；语义 Web 安全参考技术框架
中图分类号：TP309.2 **文献标识码：**A **文章编号：**1000-436X(2012)Z2-0045-15

Study on semantic Web security

CHEN De-yan[1,2], ZHAO Hong[1], ZHANG Xia[1,2], ZHAO Li-jun[2], PING An[2]
(1. National Engineering Research Center for Computer Software, Northeastern University, Shenyang 110179, China;
2. State Key Laboratory of Software Architecture, Neusoft Corporation, Shenyang 110179, China)

Abstract: Semantic Web technologies themselves as well as the web of data constituted by semantic Web applications on the Internet might introduce new security, privacy and trust issues as well as requirements. So, security issues faced by the Semantic Web and its security status were analyzed in detail. On this basis, the semantic Web security reference framework was proposed. Finally, urgent needs about key technologies, policy specifications and safety standards of semantic Web were pointed out based on the semantic Web security reference framework.
Key words: semantic Web; ontology; semantic Web security; semantic Web security reference framework

1 引言

今天的 Web 内容，大多数只适合人类消费[1]。除了基于关键字的搜索引擎以外，Web 应用活动很少有软件工具的支持。基于关键字的搜索引擎虽然对 Web 应用具有很大的价值，但仍然存在很多不足。比如：返回太多结果（其中可能包含了相关的页面）、返回很少或者没有结果、搜索结果高度依赖搜索的关键字（需要发起多次搜索以查找相关内容，不能返回语义上相似的结果）、结果是单个 Web 页（必须人工从各个 Web 页抽取相关内容并组合到一起）、搜索结果很难被其他软件工具所使用等。

今天的 Web 关键问题是：Web 内容的含义因为缺乏语义标注，并不是机器可访问的。虽然有一些分词工具可以实现分词、词性标注等功能，但对于解释内容的含义或为用户抽取有用的信息，其能力仍然非常有限。例如，这些工具很难区分这几句话的含义：“我很想去”、“我不是不想去”、“你可能认为我很想去”。

为改善这样的状况，一种解决方法是仍然使用像今天一样展示的 Web 内容，并基于人工智能和计算语言学开发日益复杂的技术。到目前为止，尽管

收稿日期：2012-10-23
基金项目：国家重点基础研究发展计划（“973”计划）基金资助项目（2012CB724107）；电子信息产业发展基金资助项目
Foundation Items: The National Basic Research Program of China (973 Program) (2012CB724107); The Electronic Information Industry Development Fund Project

这种方法取得了一些进步，但这个任务仍然面临着太多的困难。另一种方法是采用一种更容易被计算机处理的形式表述Web内容，并使用智能技术利用这些表述，即语义Web[2]方法。语义Web并不是和现有的万维网（WWW, world wide web）对等的竞争物，而是现有Web的一种逐步演化。

语义Web通过构建数据或实体的多种上下文关系赋予数据或实体丰富的语义，语义Web技术解决的2个关键问题是语义表示和语义推理，实现机器自动理解Web页的语义并实现自动的语义交互。语义Web技术的基础是可扩展标记语言（XML, extensible markup language），XML在词法和概念上有一定的表示能力，但是在描述资源和资源之间的语义关系以及推理能力方面仍然不足。为了解决这个问题，语义Web先后提供了多种知识表示语言，每一种都对表达和推理能力进行了进一步的扩展，并且允许用户根据语义程序具体所需的语义量或语义描述和推理能力的平衡来选择相应的表示方式。

随着语义Web技术和相关标准的发布和成熟，支持语义Web应用的工具也大量出现，例如语义Web编程框架（Jena、Sesame等）、本体编辑器（protege）、本体存储产品（Oracle 11g、OWLIM、Parliament、Stardog、Jena语义Web框架提供的TDB等）、本体推理机（Pellet、FaCT++、HermiT等）、规则引擎（Pellet、Jess、Drools等）等。而且，越来越多基于语义Web技术的应用开始涌现，例如语义Wiki、FOAF[3]（friend of a friend）、语义搜索、语义Web services[4]（缩写为SWS）等。设想在语义Web技术广泛应用的背景下，基于语义Web技术所构建的数据互联网络，用户将可以更加迅速、准确地找到几乎全部想要的信息。这些信息既可以是物理存在的实体（比如网络可访问的电子文档、图像、服务或者网络不可访问的人、组织、图书馆的书等），也可以是物理上不存在的抽象概念（比如“创建者”的概念）。

语义表示和推理能力的引入，本身并不是要揭示内容和隐私，但会被一些数据分析和挖掘工具所利用，而导致敏感信息或隐私数据的泄露。由于语义Web技术组件的引入和应用实现架构的变化，使得基于语义Web技术的应用需要解决新的安全、隐私、信任等问题和需求。而这些安全问题的尽快解决，反过来又会推动语义Web技术的广泛应用和发展。

本文首先分析语义Web所面临的安全问题和安全现状，提出未来语义Web的安全技术框架及重要的科研方向，以期为未来语义Web安全的研究、成功应用和产业发展做出有益的探索。

2 语义Web技术介绍

语义Web是当前Web的未来愿景，也被称为Web 3.0。语义Web通过采用形式化的、机器可处理的语义Web语言来标注Web资源的语义，最终让机器代替人做更多的工作，实现Internet上不同Web资源的自动发现、自动集成、共享和重用，并支持通过互联网的信任交互。

2.1 语义Web技术层次架构

Tim Berners-Lee在2000年的XML大会上给出了语义Web的定义，“语义Web并不是一个单独的Web，而是当前Web的扩展，语义Web上的信息被赋予了正确的含义，从而使计算机和人能够更好地协作。”同时也提出了它的层次体系结构框架模型[5]，如图1所示。

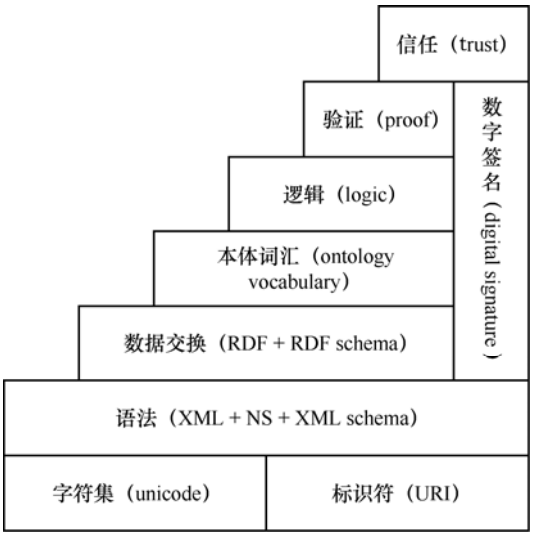

图1 语义Web层次结构

第一层是基础层。包括统一编码（unicode）和统一资源标识符（URI, uniform resource identifier）。Unicode是2个字节的全编码，可对各国文字和符号进行统一编码，使得任何语言的字符都可以被机器处理。URI可以用于唯一地标识任何物理或者抽象资源。第二层是语法层。其核心是XML及相关规范。可以使用用户自定义的词汇来编写结构化的Web文档，并使用命名空间唯一地标识元素和属性名，实现跨Web的交换。第三层是资源描述层。资

源描述框架[6]（RDF, resource description framework）是一个用于编写 Web 对象（资源）简单陈述（statements）的基本数据模型。RDF 数据模型不依赖 XML，但 RDF 具有基于 XML 的语法。RDF schema[7]（缩写为 RDFS）为将 Web 对象组织为层次结构提供了建模原语。关键原语是类（classes）和属性（properties）、子类（subclass）和子属性（subproperty）关系、定义域（domain）和值域（range）约束。第四层是本体（ontology）词汇层。本体提供了比 RDFS 具有更丰富的表达能力语言，可以表示 Web 对象之间更复杂的语义关系。当前的 Web 标准是 Web 本体语言[8]（OWL, Web ontology language）。第五~第七层是逻辑、验证和信任。用于处理信任管理和协商，即信任 Web 给笔者的信息。逻辑层在前面各层的基础上执行基于公理和规则的逻辑推理操作。验证层根据逻辑陈述进行验证，以得出结论。信任层处理信任管理和信任协商。最后，贯穿第三~第六层的是数字签名，用于支持 Web 信任机制的实现。随着对语义 Web 的深入研究，语义 Web 层次结构也在不断地发展演变和进化。

2.2　XML

XML 是一种用于定义标记、使用很普遍的元语言，它为应用数据和元数据之间的交换提供了一种统一的框架。HTML 用于描述数据及其表现形式，而 XML 被设计专门用于描述数据，相同的 XML 数据，根据需要可以表现为不同的形式。HTML 有一套固定的、预定义的标签词汇集，用于表示数据的显示格式。而 XML 允许用户通过文档类型定义（DTD，document type definition）或 XML schema 定义 XML 文档包含的合法元素、属性以及这些元素、属性的合法结构。XML 的本质是使标记对人和机器都可读。与 HTML 比较，XML 具有很多优势。

2.3　RDF

XML 没有提供任何方式表示数据的语义（含义）。例如，XML 并没有标准的方式为标签嵌套提供关联的含义，需要由应用来解释嵌套的含义[1]。XML 在提供机器可处理文档方面的能力是有限的。为了弥补 XML 的不足，语义 Web 提出了 RDF 语言来增强 XML 的语义描述能力，并为交互、搜索和编目(catalogue)提供了更好的支持。

RDF 是一种资源描述语言，它专门用于表达关于资源的元数据，比如 Web 页面的标题、作者和修改时间，文档的版权和许可信息等。资源可以是任何具有标识的事物，这些事物可以是物理上存在的实体或者物理上不存在的抽象概念，可以是 Web 可访问的或者不可访问的。RDF 用于信息需要被应用程序处理而不是仅仅显示给人观看的场合。RDF 提供了一种用于表达这一信息、并使其能在应用程序之间交换而不丧失语义的通用框架。

尽管 RDF 被称为一种资源描述语言，RDF 本质上是一个数据模型。RDF 提供了一种简单的方式来做出关于 Web 资源的陈述，陈述是语义 Web 的基本构成单元，可以说语义关系（即陈述）构成了语义 Web。RDF 基于这样的思想：用 Web 标识符（URIs）来标识事物，被描述的事物具有一些属性，而这些属性各有其值；对资源的描述可以通过对它做出指定了上述属性及值的陈述来进行。RDF 用一套特定的术语来表达陈述中的各个部分。确切地说，陈述中被描述的资源称为主体（subject），与资源关联的属性（例如：作者、创建日期、语种等）称为谓词（predicate），属性的值称为客体（object）。主体和谓词都是资源，客体可能是一个资源，也可能是一个字面值（literal）。

陈述或陈述的集合具有 3 种视图：RDF 图、三元组和 RDF/XML。组成陈述的各个部分构成了一个三元组（triple）。这使得 RDF 可以将一个或多个关于资源的简单陈述表示为一个由结点和弧组成的图，其中的结点代表资源或属性值，弧代表属性。为实现语义 Web 愿景中机器可处理的方式来表示陈述，RDF 采用了可扩展标记语言 XML。XML 被设计成允许任何人来设计它们自己的文档格式，并可用这种格式书写文档。RDF 定义了一个特殊的 XML 标记语言（称为 RDF/XML）来表示 RDF 信息和在机器间交换这些信息。XML 标签能让那些理解这些标签含义的程序正确地解释文本内容。XML 的内容和标签（除了一些特例）能够包含 unicode 字符，这就允许来自各种语言的信息可以被直接显示出来。

2.4　本体

术语“本体”最初起源于哲学领域，其关心的是客观现实的抽象本质，是对客观存在的一个系统的解释或说明。它在哲学中的定义为“对世界上客观存在物的系统描述，即存在论。”在人工智能界，最早给出本体定义的是 Neches 等，他们将本体定

义为“本体定义了组成主题领域词汇表的基本术语及其关系，以及结合这些术语和关系来定义词汇表外延的规则。”后来在信息系统、知识系统等领域，越来越多的人研究本体，并给出了许多不同的定义。其中，最著名并被引用得最为广泛的定义是由 Gruber 在 1993 年提出的“本体是概念化的明确的规范说明。”之后 Studer 在 Gruber 的基础上于 1998 年扩展了本体的概念，即 Studer 认为“本体是共享概念模型的明确的形式化规范说明。”这个定义体现了本体的 4 层含义：1) 概念模型(conceptualization)：通过抽象出客观世界中一些现象的相关概念而得到的模型，其表示的含义独立于具体的环境状态。2) 明确(explicit)：所使用的概念及使用这些概念的约束都有明确的定义。3) 形式化(formal)：本体是计算机可读的，即能被计算机处理。4) 共享(shared)：本体中体现的是共同认可的知识，反映的是相关领域中公认的概念集，它所针对的是团体而不是个体。

本体的目标是捕获相关领域的知识，提供对该领域知识的共同理解，确定该领域内共同认可的词汇，并从不同层次的形式化模式上给出这些词汇（术语）和词汇间相互关系的明确定义。本体克服了在术语上的差异（通过本体映射），实现语义互操作性。总的来说，构造本体可以实现某种程度的知识共享和重用，以及提高系统通信、互操作、可靠性的能力。

根据特定领域的需要，有多种本体分类方法。Fensel 给出了一种有趣的本体分类方法，根据本体对某个特定任务或观点的依赖级别来分类本体[9]，可以将本体分类为高层本体(或顶层本体)、领域本体、方法本体、任务本体、领域任务本体以及应用本体。有多种用于描述本体的语言，而 OWL 语言被认为是最有前途的本体语言。

2.5 OWL

RDF/RDFS 允许一些本体知识的表示，然而，RDF/RDFS 的表达能力是非常有限的。RDF/RDFS 的主要建模原语涉及到以类型化的层次组织词汇：子类和子属性关系、定义域与值域约束和类实例。RDF/RDFS 缺失许多其他有用的特性[1]，例如类的不相交性（disjointness）、类之间的布尔组合、类之间的等价、属性的等价、属性的基数（cardinality）限制、个体（实例）的同一性、属性链等。

W3C 的 Web 本体工作组为语义 Web 识别了许多特征用例，这要求比 RDF 和 RDF Schema 提供更丰富的表达能力。在美国和欧洲的许多研究小组已经识别到更强的本体建模语言的需要，这导致了一个联合倡议以定义一种更丰富的语言，称为 DAML+OIL[10]。这个名字是美国提案语言 DAML-ONT[11]和欧洲提案语言 OIL[12]的一个组合。DAML+OIL 反过来成为 W3C 的 Web 本体工作组定义 Web 本体语言（OWL）的一个起点[1]。

OWL（最新版本称为“OWL 2”）的目标是提供一种标准的、广泛接受的用于语义 Web 的本体语言。OWL 通过提供更多具有形式语义的词汇，使之在 Web 内容的机器可理解性方面要强于 XML、RDF 和 RDFS 等所能达到的程度。在设计这样的语言时，需要意识到表达能力和高效推理支持之间的平衡。语言的表达能力越丰富，推理支持变得越无效率的。为此，OWL 提供了 3 种表达能力递增的子语言（OWL Lite、OWL DL 和 OWL Full），以分别用于特定的实现者和用户团体。在表达能力和推理能力上，每个子语言都是前面语言的扩展。

在实践中，为了存储 OWL 2 本体并在工具和应用之间交换它们，需要一个具体的语法。用于 OWL 2 的主要交换语法是 RDF/XML，这是所有 OWL 2 工具必须支持的唯一语法。尽管 RDF/XML 提供了 OWL 2 工具之间的交互能力，但也可以使用其他具体的语法，例如函数式（functional-style）语法、曼彻斯特（manchester）语法、OWL/XML 语法、Turtle 语法，不同的语法适用于不同的目的。有一些工具可以实现不同语法之间的转换。

OWL 本体的典型组件包括：1）类：一个 OWL 类是一种特殊的资源，它描述了具有共同特征或在某方面相似的资源的一个集合。2）个体（individuals）：个体也是一种资源，它是类的成员，代表了该类的一个实例。3）属性：属性也是一种资源，它在描述个体的陈述中充当谓词。OWL 中有 2 种主要类型的属性：对象属性用于将一组个体和另一组个体连接起来；而数据类型属性用于将个体和文字值连接起来。在 Web 上，OWL 本体一般以文档的形式进行存储，或者以三元组的形式存储在关系数据库或 RDF 专用存储中。

2.6 规则

语义 Web 层次结构提供了多种知识表示形式，包括从 RDF 到最新版本的 OWL 等多种格式，每一

层都对表达能力进行了进一步的扩展，并且允许用户根据语义程序具体所需的语义量来采用相应的表示方式。在 OWL 2 出现之前，OWL 1 为用户提供的表达形式在数量上还很有限。解决这一问题的一种方案就是保持 OWL 推荐标准不变，而在此基础上支持规则的加入，进而扩展 RDF/OWL 数据的表达能力。OWL 2 的发布已经解决了原来 OWL 1 推荐标准中的一些缺陷。

到目前为止已经出现了多种规则语言和引擎，如 RuleML（rule markup language）、DAML（DARPA agent 标记语言）、语义 Web 规则语言[13]（SWRL, semantic Web rule language）、WRL（Web 规则语言）、SWSL（语义 Web 服务语言）、F-Logic、Prolog、Jess 规则语言和规则引擎、Jena 规则语言和规则引擎、Drools 规则语言和规则引擎等。它们已经以形形色色的方式应用到了语义 Web 之中。

目前语义 Web 还没有标准的规则语言。语义 Web 规则事实上的标准语言是 SWRL，SWRL 得到了强有力的公众认同和工具支持，有广泛的用户基础，而且它基于的是本领域中一些最受人尊敬的科研人员的工作成果[14]。SWRL 是一种具有丰富表达力的基于 OWL 的规则语言。SWRL 允许用户编写可以用 OWL 概念术语表示的规则，以提供比孤立的 OWL 更强大的演绎推理能力。在语义上，SWRL 建立在和 OWL 相同的描述逻辑基础之上，并且当进行推理时，提供了类似的、健壮的正式保证。SWRL 是基于 OWL 的 OWL DL 和 OWL Lite 子语言和 RuleML 的 Unary/Binary Datalog RuleML 子语言（基于 Horn 子句）的一种组合。SWRL 的目标是提供 OWL 1 所不支持的表达能力，同时保持与 OWL 语法、语义和理论模型的兼容性。

2.7　SPARQL

虽然 RDF 支持基于 XML 的 RDF/XML 存储和交换语法，但现有的 XML 查询语言（Xpath）并不适合查询 RDF 数据。因为 XML 位于比 RDF 更低的抽象层次，如果基于 XML 的查询语言来查询 RDF 数据，将会是很复杂的问题。对于某个 RDF 陈述，如果采用 XML 来进行表示，将有多种词法格式，而它们在语义上是等价的。如果采用 Xpath 查询，针对这些不同的词法表示，将需要编写不同的 Xpath 查询语句。为了解决这个问题，一种更好的方式是寻找一种可以在 RDF 层次编写查询语句的语言。这种查询语言必须理解 RDF，也就是说，它不仅理解语法，而且理解 RDF 的数据模型和 RDF 词汇的语义。这样的查询语言就是 W3C 推荐标准 SPARQL[15]。几乎所有的 RDF 查询工具都提供了对 SPARQL 查询语义的支持，尽管存在出现更早、更成熟且比 SPARQL 具有更丰富的表达特性集的查询语言（例如，SeRQL 和 RQL 语言），但支持它们的工具却很少，阻碍了交互能力。SPARQL 基于 RDF 三元组图模式来匹配查询。

3　语义 Web 的安全问题

语义 Web 的安全问题涉及语义 Web 层次架构中的各个层次以及信息交互、数据集成、隐私、信任等方面。除了技术层面的安全问题以外，还涉及到相关的安全标准规范的建立和完善。

3.1　语义 Web 组件本身的安全问题

3.1.1　XML 的安全问题

XML 的出现，弥补了 HTML 的很多缺陷，并迅速获得了广泛的应用。语义 Web 层次架构也基于 XML 之上进行构建，构建于 XML 之上的语义 Web 语言进一步增强语义表示与推理能力。

由于 XML 的广泛应用，XML 的安全问题也很快获得了学术界、标准组织以及企业界的关注，并获得了广泛的研究和应用。XML 的安全问题如表 1 所示。

3.1.2　RDF 的安全问题

RDF 的安全同样涉及到 RDF 文档的安全、RDF 数据库的安全和 RDF 的安全标准。RDF 的安全问题如表 2 所示。

3.1.3　本体的安全问题

在确保 XML 和 RDF 安全的基础上，同样要确保本体的安全。OWL 本体陈述和实例陈述往往是放在同一个本体中的，但也可以分散在 OWL 文件中。本体的安全问题如表 3 所示。

3.1.4　规则的安全问题

规则的安全问题涉及到规则的存储、规则的访问、基于规则的推理和规则安全标准等，如表 4 所示。

3.2　安全的数据集成

对于信息互操作性，包括处理数据异质性和策略异质性，本体已经成为了公共的实践。下层语义为这些数据提供推理支持，可以产生意外的结果。OWL 相应地提供了用于类、属性和个体之间映射（对准）的一些属性，如表 5 所示。

表 1　XML 的安全问题

XML DTD, XML schema 的安全	由于 DTD 和 XML schema 包含了 XML 文档所使用的元素以及结构信息，可以用于解析遵循该定义的所有 XML 文档，所以其安全性非常重要，需要考虑是否允许对其进行访问
XML 文档的安全	1）构成 XML 文档的各个组件的安全，包括 XML 元素、XML 属性、XML 命名空间。例如，不同 XML 元素是否具有不同的安全级别；对某个元素的访问授权是否可以传播到其子元素等 2）对整个 XML 文档或文档某部分的访问控制。仅暴露具有访问权限的文档视图 3）XML 文档的安全发布。用于第三方发布时如何确保文档的真实性和完整性
XML 数据库的安全	1）XML 数据的存储安全。例如进行加密存储 2）XML 数据的安全访问。例如，如何基于安全策略对非授权的访问操作进行过滤或者修改 3）XML 数据的安全集成。来自多个数据源的 XML 数据如何进行安全集成
XML 的安全标准	W3C(world wide Web consortium)和 OASIS (organization for the advancement of structured information standards)分别开发了一些与 XML 安全相关的标准规范。W3C 已经开发了与安全相关的 3 个主要标准：XML 加密[16]（XML encryption）、XML 密钥管理[17]（XML key management）以及 XML 签名[18]（XML signature）。OASIS 是一个推动 Web services 安全的非盈利国际标准组织。OASIS 提供的 2 个主要标准是安全断言标记语言[19]（SAML, security assertion markup language）和可扩展访问控制标记语言[20] (XACML, extensible access control markup language)

表 2　RDF 的安全问题

RDFS 的安全	不像 XML schema 约束了 XML 文档的结构，由于 RDF schema 定义了在 RDF 数据模型中使用的词汇以及对象之间的关系，例如子类关系、子属性关系、属性的定义域和值域。那么对关联关系（属性）的保护就涉及到是否允许对整个 RDF schema 或者其某个部分的访问，比如对各个部分进行安全分级
RDF 文档的安全	1）语法安全。为确保 RDF 文档的安全，首先要从语法层面确保 XML 的安全 2）语义安全。从语义层面确保 RDF 文档的安全。这些问题涉及概念资源、属性、陈述的安全蕴含（implications）和安全约束。例如，允许对陈述中资源（不包括属性，因为在语义 Web 中属性也是一种特定类型的资源）的访问，而不允许对资源之间关联关系（即属性）的访问，反之亦然。或仅允许对部分陈述的访问 3）对整个 RDF 文档或文档某部分的访问控制。仅暴露具有访问权限的文档视图 4）RDF 文档的安全发布。用于第三方发布时如何确保文档的真实性和完整性
RDF 数据库的安全	1）RDF 数据的存储安全。例如进行加密存储 2）RDF 数据的安全访问。例如，如何基于安全策略对非授权的访问操作进行过滤或者修改 3）RDF 数据的安全集成。来自多个数据源的 RDF 数据如何进行安全集成
授权传播问题	在某个概念或关系上的授权，基于语义关系和推理，可能会传播到相关的概念或者关系，导致隐含的非授权的访问
RDF 的安全标准	为了实现统一的可交互安全措施，必须制定相应的安全标准。但目前还没有用于 RDF 的安全标准

表 3　本体的安全问题

OWL 本体（模式）陈述的安全	OWL 本体描述了领域中的概念及其语义关系，用于标记领域中的实例数据。尤其当涉及到用于描述个人基本信息或个人健康档案的本体陈述时，需要考虑是否允许对整个本体或本体的某一部分的访问
OWL 实例陈述的安全	实例陈述保存了符合某个领域的本体陈述的真实实例数据，其安全性更加重要，尤其当涉及到敏感或个人隐私数据时。需要考虑是否允许对整个实例或部分实例陈述的访问
OWL 数据库的安全	1）OWL 数据的存储安全。例如进行加密存储 2）OWL 数据的安全访问。例如，如何基于安全策略对非授权的访问操作进行过滤或者修改 3）OWL 数据的安全集成。来自多个数据源的 OWL 数据如何进行安全集成
本体推理安全	基于本体中的显式陈述和 OWL 推理工具，可以推理出隐式陈述，而这些隐式陈述可能涉及到敏感或者隐私信息
授权传播问题	在某个概念或关系上的授权，基于语义关系和推理，可能会传播到相关的概念或关系，导致隐含的非授权访问
OWL 的安全标准	为了实现统一的可交互安全措施，必须制定相应的安全标准。但目前还没有用于 OWL 的安全标准

表 4　规则的安全问题

规则的安全存储	基于 OWL 本体的 SWRL 规则可以与 OWL 本体保存在一起（以 RDF/XML 语法），或者保存在单独的规则库中。这样，为保证规则的安全，首先需要保证 OWL 本体的安全
规则的安全访问	在使用 SWRL 规则表达安全策略时，更需要确保规则的访问安全，只有具有授权权限的用户才可以制定并维护规则，而且不同授权权限的用户只能制定或维护不同的授权规则
由规则推理引发的安全与隐私问题	与语义 Web 相关的隐私问题，除了前面提到的在本体集成时可能涉及到的隐私问题外，还涉及到由语义推理所引发的隐私泄露问题。规则引擎基于本体中陈述的显性事实/知识和语义 Web 规则（例如 SWRL 规则）推理出隐性事实/知识。而这些推理得到的信息可能是高度敏感或者私有的。例如，通过语义推理，如果将患者的名字和其健康记录关联到了一起，将导致隐私泄露问题
规则的安全标准	为了实现统一的可交互安全措施，必须制定相应的安全标准。但目前还没有用于语义 Web 规则的安全标准

表 5　　OWL 本体映射属性

分类	属　性	含　义
类和属性之间的等价关系	owl:equivalentClass, owl:equivalentProperty	当笔者要把一些本体组合在一起作为另一个新的本体的一部分时，能说明在一个本体中的某个类或者属性与另一个本体中的某个类或者属性是等价的
个体间的同一性	owl:sameAs	描述个体之间相同的机制，与描述类之间的相同机制类似，仅仅只要将 2 个个体声明成一致就可以
不同的个体	owl:differentFrom owl:AllDifferent owl:distinctMembers	这一机制提供了与 sameAs 相反的效果

owl:sameAs 表达等价的能力，可被用来表达表面上不同的个体实际上是相同的。另外，owl:InverseFunctionalProperty 也可被用来连接个体。例如，如果一个属性“SocialSecurityNumber”是一个 owl:InverseFunctionalProperty 属性，那么 2 个分开的个体如果具有相同的 SocialSecurity Number 属性，则可被推理出是相同的个体。当个体被这样确定为相同时，来自异源的关于这些个体的信息可以被合并。这种聚合可被用来得出不可直接从单源获得的事实。语义 Web 连接来自多源信息的能力是一个理想的、强大的特性，它可被用在许多应用中。但是合并来自多源数据的能力，加上 OWL 的推理能力，可能会导致信息被误用，最终导致隐私泄露问题[21]。例如，FOAF 项目使用 RDF 定义了一组词汇（本体），用于集成来自各个人、他们的朋友以及他们的朋友的朋友的信息（个人基本信息、好友关系等）。在这种情况下，很有可能被收集并利用到某人的隐私或者不愿意公开的信息。即使个人没有明确提供基于 FOAF 词汇的 RDF 文件，但其他好友仍有可能提供关于某个人不同方面的信息，并可以基于 FOAF 工具实现关于这个人的信息自动聚合。

3.3　信任问题

关于信任，现代汉语的解释是“相信而敢于托付”、“相信并加以任用”。基于对某个人的信任，由此人产生的数据也会被分配一个信任值。此值取决于该人是否可以保守秘密或开展安全活动等。如果数据来源于一个值得信赖的人或者源（例如文件或数据库），那么这个数据会被分配一个较高的信任值。信任的对象可以是人、一组人、一个组织、一个 Web 站点、数据、代表人执行活动的进程等。

信任管理是关于管理一个人或组对另一个人或组或者其他信任对象的信任。也就是说，即使一个人有权访问数据，笔者能信任这个人并将数据发送给他/她吗？该用户可能有许可或拥有凭据，但他/她不一定值得信任。信任依赖于用户的行为。如果用户背叛了机密性或执行了一些不适当的活动，语义 Web 不可能信任这个用户。语义 Web 既面临传统 Web 环境相同的信任问题，也面临一些新的信任问题，详细如表 6 所示。

3.4　隐私问题

隐私的概念在不同国家、文化和管辖范围间（有时是在这些范围之内）差别很大，不同国家或不同区域对公民的隐私保护有不同的法律规定和要求。什么是隐私，一般的概念是：某个人应当决定应该释放关于他/她的什么个人信息。在数据分析和数据挖掘的工具以及 WWW 出现以前，这样的定义是好的。而有了这些工具之后，对于某人来说，获取另一个人的隐私信息可能变得轻而易举。

同样，在语义 Web 环境下的隐私问题，既包含传统 Web 环境下的隐私问题，也有其新的隐私考虑，详细如表 7 所示。

4　语义 Web 安全现状

语义 Web 的发展虽然取得了很大的进步，但是，语义 Web 仍面临很多技术挑战，比如大规模推理和存储技术、本体构建、本体演化和变更、本体映射、语义 Web 安全等，尤其必须首先解决语义 Web 的安全问题。

4.1　各国政府对语义 Web 安全的关注

语义 Web 的发展受到许多行业的推动，并且政府也做出了极大的投资。美国政府建立了 DARPA（美国国防高级研究计划署）代理标记语言（DAML, DARPA agent markup language）项目[22]，DAML 努力的目标是开发一种语言和相关的工具来促进语义 Web 的实现。DAML 语言作为 XML 和 RDF 的一个扩展，其释放版本 DAML+OIL 提供了丰富的构造以创建本体和标记信息，使得使用该语言描述的内容是机器可读并且可理解的。DAML 的设计目标是在描述对象和对象之间关系上面具有

表 6　语义 Web 的信任问题

如何表示信任	对信任的表示目前主要有 3 种方法[4]是：基于凭据（credentials）、基于声誉（reputation）以及某个人为数据赋予的机密性值（confidence value）。那么，在语义 Web 环境下，如何定义和表示信任
如何协商并建立信任	信任管理的一个重要方面就是信任协商。双方可以彼此协商信任值以及在他们之间共享的数据。目前主要有基于凭据和基于声誉的信任协商方式，那么在语义 Web 环境下如何在各方之间协商信任并共享和集成数据
如何管理信任	一旦信任值被分配以后，在动态分布式语义 Web 环境下，如何动态维护信任对象的信任度的大小
能否部分地信任某人	比如，能否在 50%的时间信任 Mr. John，而在 70%的时间信任 Mrs. Jane
在部分信任的情况下，能否共享信息	如果可以部分地信任某人，是否可以共享信息给他/她以及是否可以信任他/她共享的数据
能否信任来自不信任源的数据	也即，如何信任语义 Web 给我们的信息。在语义 Web 环境下，基于语义 Web 技术所描述的陈述（例如，“A”和“B”是好朋友）分布于互联网上的各个角落，它们由不同的人所创建，那么，如何信任这些陈述的真实性或者说正确性
能否信任包含不信任陈述的推理结果	如果陈述本身不是可信任的，那么基于这些陈述的推理结论或者决策也可能是不可信任的
如何根据信任值的大小分配权限	如果对信任可以进行定量表示，那么不同的信任值和权限之间如何进行自动化的映射呢
信任规则如何表示	例如，有这样的规则定义：如果 B 共享 C 给 A，A 可以共享数据 D 给 B。仅仅如果 B 不共享数据 D 给 F，A 才可以共享这个数据 D 给 B。人们可以定义许多这样的规则 在语义 Web 环境下，资源的语义描述和信任规则的定义分布于多个数据源，那么如何实现信任规则语义的一致性
信任传递问题	如果 A 信任 B，B 信任 C，A 能否信任 C 即使 A 不信任 C，但 B 可能信任 C，他可能共享他的数据给 C。也就是说，C 对 A 来说是不值得信任的，但他对 B 来说又是值得信任的
信任的语义表示	如何将信任的定义、信任协商和信任管理的语义加入语义 Web 的技术组件之中

表 7　语义 Web 的隐私问题

隐私数据和需求的识别	在语义 Web 环境下，隐私数据位于语义上下文之中，隐私数据和需求的识别相应地需要考虑其语义上下文，而不能孤立地标识它们
隐私策略如何表示	隐私数据位于语义上下文之中，相应地隐私策略也应该支持位于本体中的语义构造，以准确地表达用户的隐私约束
本体集成的隐私问题	参见 3.2 节
由推理引发的隐私问题	参见 3.1.4 节
由语义 Web 数据挖掘引发的隐私问题	语义 Web 挖掘有 3 个方面的含义：一个是挖掘 Web 上使用语义 Web 技术（诸如 XML、RDF 和 OWL）表示的数据。另一个方面是挖掘 XML 和 RDF 文档，而没有泄露实际的数据，但分发关联关系和趋势（correlations and trends）。还有第三个方面，也就是说使用本体帮助挖掘过程。例如，数据挖掘工具可能需要澄清某个 Web 页的含义。这里，使用 OWL 表示的本体可以用于澄清概念，并辅助挖掘过程。需要采用隐私保护的语义 Web 数据挖掘，其目标是隐藏隐私数据，诸如某个人的疾病，同时分发一般的趋势和关联。也就是说，可以分发这样的信息：“生活在加利福尼亚的人们更易于得哮喘”，而没有分发这个事实：“John 有哮喘”[4]
隐私安全标准	为了实现统一的可交互的安全措施，必须制定相应的安全标准。目前还没有用于语义 Web 的隐私安全标准

比 XML 更强的能力，它可以表达语义，可以在网络站点之间创建更高的协同级别。作为美国国防部的核心研究开发署，DARPA 在创建因特网和许多技术方面做出了贡献。DAML 项目正式开始于 2000 年 8 月，包括一个集成承包商和 22 个技术开发团队。其中，有几个小组专门研究语义 Web 的机密性、完整性、信任问题，同时研究采用 DAML+OIL 语言来定义安全策略。

语义 Web 及安全、隐私和信任技术也是欧盟第 6 框架计划[23]（FP6, European Union’s sixth framework programme for research and technological development）和 FP7[24]（欧盟第 7 框架计划）的关键行动路线之一。FP6 开始于 2002 年，于 2006 年结束，其中基于知识的技术、信任和安全被应用于多个主要研究领域之中。FP7 开始于 2007 年，将持续到 2013 年。在在 FP7 中，信息和通信技术、安全是其 10 个主要研究中的 2 个研究领域。

英国前首相戈登・布朗 (BROWN G)曾宣布投入 3 000 万英镑用于语义 Web、新兴 Web 和 Internet 前沿技术的研究[25]，以促进英国在社会经济处于世界领先的位置。这也是布朗首相第一次在同一句话中同时提到“语义”和“Web”。这项研究由语义 Web 的提出者 Tim Berners-Lee 先生参与领导。随着语义 Web 技术的进步以及语义 Web 在各个领域应用中的不断出现，相信语义 Web 技术以及语义 Web 的安全问题将吸引更多政府组织的关注。

4.2　国内外安全标准组织及其进展

W3C 和 OASIS 分别开发了一些与 XML 安全相关的标准规范。W3C 已经开发了与安全相关的 3 个主要标准：XML 加密、XML 密钥管理以及 XML 签名。XML 加密是一个用于加密/解密数字内容（包括 XML 文档）的过程，由 XML 加密工作组开发。XML 密钥管理是一个用于客户端从 Web 服务端获取密钥信息（例如密钥、证书等）的协议，由 XML 密钥管理工作组开发。XML 签名是一个 XML 兼容的语法用于表示 Web 资源和协议消息部分（可以使用 URI 标识的任何事物）以及用于计算和验证这个签名的过程，由 XML 签名工作组开发。

OASIS 是一个推动 Web services 安全的非盈利国际标准组织。OASIS 提供的 2 个主要标准是 SAML 和 XACML。SAML 是一个用于在不同的安全域(security domain)之间交换认证和授权信息的 XML 框架。在 SAML 标准中分别定义了身份提供者 (identity provider) 和服务提供者 (service provider)，这两者构成了前面所说的不同安全域。XACML 是一种用于决定请求/响应的通用访问控制策略语言和执行授权策略框架，它在传统的分布式环境中被广泛用于访问控制策略的执行。在典型的访问控制框架中，有策略执行点(PEP, policy enforcement point)和策略决定点(PDP, policy decision point)。PEP 用于表达请求和执行访问控制决定。PDP 从 PEP 处接受请求，评估适用于该请求的策略，并将授权决定返回给 PEP。

隐私首选项平台[26]（P3P, platform for privacy preferences）是由 W3C 推荐的并由几家主要的公司（Microsoft、IBM、HP、NEC、Nokia 和 NCR）参与制定一个新兴的行业规范，旨在为网上冲浪的 Internet 用户提供隐私保护。现在有越来越多的网站在消费者访问时，都会收集一些用户信息。制定 P3P 标准的出发点就是为了减轻消费者因网站收集个人信息所引发的对于隐私权可能受到侵犯的忧虑。如果 Web 站点实现了 P3P，当用户进入该 Web 站点时，P3P 使得 Web 站点可以以一种标准的格式表示它们的隐私保护策略（包括了网站需要用户提供的个人隐私信息以及对这些信息所做的处理），这个策略的格式可以被用户代理自动获取和理解。P3P 的初始版本使用 RDF 指定策略，而最近的版本已经迁移到 XML。

W3C 先后推出了 RDF、OWL、SPARQL 等语义 Web 相关的规范，但并未推出与这些语义 Web 技术相对应的安全规范，即目前还没有专门用于语义 Web 环境的安全策略规范。

4.3　国内外语义 Web 安全技术现状

随着对语义 Web 应用或支持的探索，对于语义 Web 的安全，也引起了学术界和 IT 产业界的关注。例如，针对语义 Web 技术组件本身的安全问题和解决方案，Thuraisingham 进行了详细分析和说明[4]，例如，针对 XML、RDF、OWL 和语义 Web 规则中要保护的元素或属性，定义不同的安全保护等级、授权的访问主体以及允许执行的访问操作。

针对语义 Web 的访问控制、隐私保护和信任协商模型以及策略规范，很多研究者分别从不同的角度和需求出发，提出了各种不同的模型。基于语义 Web 语言的描述和推理能力，Thuraisingham 也尝试了使用语义 Web 语言来作为安全、隐私、信任和完整性策略规范，并提出了访问控制和隐私控制器架构[4]。针对 SWS（语义 Web 服务）的安全问题，Thuraisingham 描述了 SWS 的定义和安全问题，提出了使用语义 Web 语言增强 WSDL 的语义描述能力，并作为安全策略规范，实现语义 Web 服务基于语义的访问控制[4]。其他研究方向和成果请参见第 6 节。

针对语义 Web 数据的存储安全，目前，有很多可用于语义数据的存储产品，既有专用于语义数据的存储产品，也有支持语义数据存储的传统数据库产品；既有开源的产品，也有需要商业授权的产品。这些语义数据存储产品，它们大多都提供了对存储语义数据安全的支持。例如，Oracle 数据库提供了对语义 Web 技术很好的支持，比如语义数据抽取和编辑工具、RDF/OWL 数据管理、SQL & SPARQL 查询、语义推理、语义规则、语义索引、版本管理等，同时，通过 VPD（virtual private database）和 OLS（oracle label security）技术提供对存储语义数据的安全管理。VPD 通过本体构造提供对存储语义数据的访问控制策略。OLS 为语义数据提供安全分级标签，标签可以定义在三元组级或者资源级（主体、谓词等），以控制对三元组或资源的访问。

5　语义 Web 安全参考技术框架建议

解决语义 Web 安全问题，首先需要根据安全威胁和安全需求，建立语义 Web 的安全参考技术框架，并对参考框架涉及到的各个关键技术展开深入的研究。本节抛砖引玉，提出了一个语义 Web 安全参考技术框架建议，如图 2 所示。

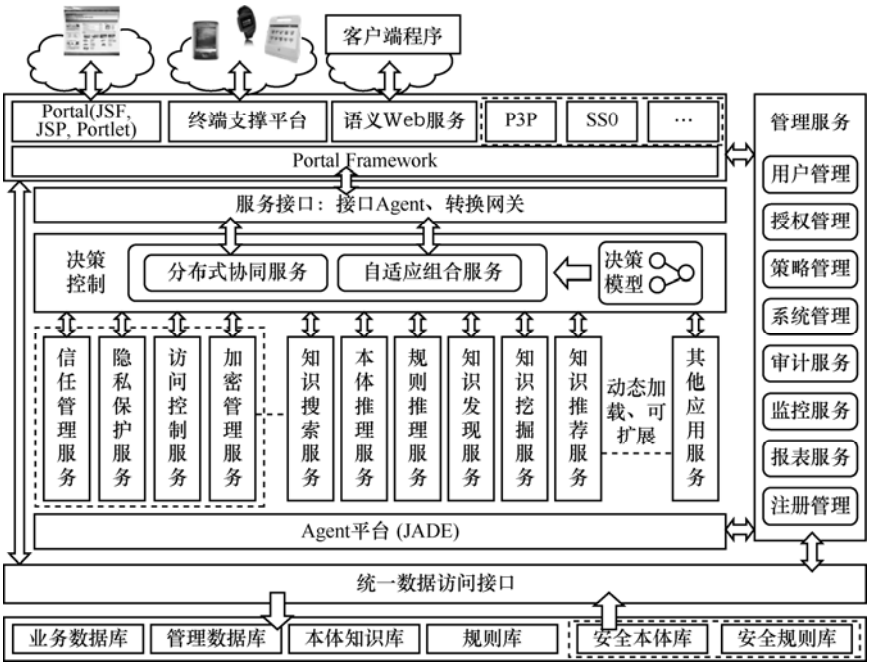

图 2　语义 Web 安全参考技术框架

整个参考技术框架分成了 4 层，自底向上分别为：数据和知识层、数据抽象层（数据统一访问接口）、平台层和展示层。

1) 数据和知识层。提供整个系统所需要的业务数据、管理数据、语义知识和语义规则。同时，为平台层的安全相关服务提供本体构造和规则。也包括存储产品提供或者定制化的其他数据存储安全机制。

2) 数据抽象层。数据抽象层统一了下层关系数据库、本体库、规则库或其他专用存储的访问接口，为平台层、展示层以及相关管理服务提供了统一的数据访问接口。根据需要，可以提供相关的安全访问接口。

3) 平台层。为了使平台服务组件高度可重用，并且，基于特定应用场景，可以很容易地修改、适配或个性化系统的行为特征，平台层采用了具有自适应特性的 Agent 平台（JADE）。并且将安全、本体查询和推理相关服务封装为可重用的 Agent 服务。比如，结合隐私保护和知识挖掘服务，可以提供隐私保护语义 Web 数据挖掘服务。平台层安全相关的服务，根据需要可以基于数据和知识层的本体构造和规则描述相应的安全、隐私、信任、加密或完整性策略。

4) 展示层。在语义 Web 分布式环境下，为了提供泛在接入服务，展示层基于 Portal 框架提供基于 Web 浏览器、智能终端和 Web services 的多种接入渠道。同时提供相应的 P3P 隐私保护、单点登录等安全服务。

参考技术框架还包括一系列管理服务，其中，包括与安全相关的管理服务，例如用户管理、授权管理、策略管理、审计服务、监控服务等。

6　语义 Web 安全关键技术研究

关于语义 Web 安全的研究仍然是一个新兴的研究领域，但已经开始引起学术界、标准化组织以及政府的关注。语义 Web 技术的发展和应用，必须首先要解决语义 Web 面临的安全问题和新的安全需求。通过对语义 Web 安全的调查和分析，本节说明语义 Web 安全需要研究的一些关键技术。

6.1　语义 Web 访问控制

在语义 Web 应用的数据模型中，基于孤立实体而忽略实体之间语义关系的授权决策可能导致未授权用户的非法推理或不完整的授权。而传统的访问控制模型（例如 DAC、MAC、RBAC 等）由于没有考虑语义 Web 数据模型中丰富语义关系，无法控制一些可推理的隐私敏感信息的泄露，并且无法验证策略执行的一致性。

针对 XML 文档的访问控制，Bertino 等提出并开发了一个用于 XML 文档的访问控制模型 Author-X，该模型主要集中在访问控制策略和传播

策略[29]。该模型使用 XML 作为策略规范，策略规范包含哪些用户可以访问文档的哪些部分信息。同时，针对 XML 文档的安全发布，Bertino 等提出了当 XML 文档用于第三方发布时确保文档的真实性和完整性的技术[30]。XML 文档在 Web 上的安全发布的思想是：访问主体向文档所有者请求对文档的访问，并从文档所有者获得访问授权；然后文档所有者将文档发布给不信任的第三方发布商；最后第三方发布商基于文档所有者给访问主体的授权，将授权的文档视图给访问主体。

针对语义 Web 应用带来新的安全需求，Javanmardi 等提出了一个用于保护语义 Web 资源的基于语义的访问控制模型[31]（SBAC, semantic-based access control model）。像其他访问控制系统一样，SBAC 基于 3 个域进行访问授权：主体、客体和行为，但 SBAC 充分考虑到了这 3 个域中实体之间的语义关系。SBAC 包含 3 个基本的组件：本体库、授权库和操作。针对授权的 3 个域，对应 3 个本体：主体本体、客体本体和行为本体。授权域是一套形式为（s, o, ±a）的授权规则，s 表示代表主体的有效凭据，o 表示客体，a 表示行为。预定义的访问授权可以以这种形式保存在授权库中。对于来自访问主体的访问请求，将基于请求的授权、本体库中的语义关系和授权库中的预定义访问授权规则进行语义推理，以得出没有明确给出但隐含的授权规则。为了防止在 3 个访问控制域上的基于语义推理的授权传播，SBAC 允许定义优先级更高的例外规则。操作是指可以在授权库上执行的操作，包括做出授权（添加规则）和吊销授权（删除规则）。SBAC 基于 OWL 本体语言描述主体之间、客体之间以及行为之间的语义关系；为了增强授权规则的表达能力，SBAC 采用了 SWRL 规则语言。

针对 SNS（social network(ing) system）系统隐私设置存在的问题，Masoumzadeh 等提出了一种基于本体的访问控制模型（OSNAC, ontology-based social network access control）[32,33]。该模型基于本体捕获社会化网络系统中复杂的实体及实体间错综复杂的语义关系，对应建立社会化网络系统本体（SNO, social network systems ontology）进行建模，并识别出安全和隐私保护需求，然后针对要保护的语义关系建立对应的访问控制本体（ACO, access control ontology）和访问控制策略规则。该模型直接将语义关系作为被保护对象，其策略框架同时支持用户级和系统级的策略，针对要保护的语义关系，支持多种授权类型。对于来自访问主体的访问请求，将基于 SNO、ACO 和访问控制策略规则进行规则推理，以授权或拒绝访问主体的访问请求。OSNAC 使用 OWL 表达在 SNS 中需要保护的知识及一些访问控制决策信息；使用 SWRL 指定访问控制策略规则；使用 SPARQL 语言和 SPARQL 协议执行访问控制。

K4CARE 项目[34]提出了一种基于用户 Profile 的访问控制模型。对系统中需要的各种可能角色定义不同的访问 Profile，每个 Profile 包含不同的原子权限（atomic rights）集合，大多数原子权限有有效域。Profile 有有效期，到期后需要更新。用户 Profile 基于 APO（APO, actor profile ontology）本体[35]构建，APO 包含了一套概念层次结构，第一层概念包括实体（entities）、服务（services）、行为（actions）、过程（procedures）、SDA（state decision action）、护理单元元素（care unit elements）和文档（documents）。实体包含角色（actor，例如患者、医生、护士等）或者不同角色为提供某个服务所组成的组（group）。不同的实体可以执行不同的行为，通过属性 doesAction 来描述。不同的角色也有一套不同的文档访问权限（读或写），例如属性 readsDocument 和 readsSometimesDocument 指 Actor 是否能使用某个文档。readsDocument 表示某个 Actor 总是能读的文档列表，无论它是否与文档中引用的患者相关。而 readsSometimesDocument 表示仅仅 Actor 与文档中引用的患者有关时才能读取的文档列表。属性 writesDocument 和 writes Sometimes Document 的含义类似。属性 initiates Service 表示实体能发起的服务。

Carminati 等基于 RDF 提出了一种安全架构[36]，该架构的主要目的是想减少安全管理员的负担，或者说自动化或半自动化授权策略的生成。在该架构中，使用 RDF 作为策略规范，描述一般领域场景及安全需求，即安全增强的 RDF 描述，并通过高层策略生成器生成高层授权策略。然后对于特定场景的 RDF 描述，通过授权继承算法自动生成授权策略。不过，这要求安全增强 RDF 描述的一般领域场景具有不同特定场景的相似特征或者说高层语义，例如，相似的资源类型、不同资源类型之间相似的语义关系以及相似的安全需求。

6.2 语义 Web 隐私保护

信息共享与隐私保护的目标是共享数据，同时保护个人身份信息，并确保信息的使用与信息收集的目的相一致。隐私保护的核心是隐私控制，即信息的拥有者是否有权完全控制个人信息（全部或部分）的收集、访问和共享，并可以要求停止对个人信息的处理。随着信息化程度的不断提高，越来越多的组织和个人开始担心安全和隐私问题，安全和隐私也成为信息化时代需要首先解决的关键问题。

Nivargi 基于语义 Web 技术提出了一种通用的隐私策略模型[37]（GPPM, generic privacy policy model），企图解决在当前数据隐私模型中的一些不足。该模型的领域和策略本体采用 OWL 语义进行描述，选择 OWL 的理由是它的描述能力可以以一种精确的方式捕捉相关的语义信息并辅助推理过程。而相应的策略语言和策略引擎选择 Rei 策略语言和 Rei 策略引擎。

针对由推理引发的隐私问题，Thuraisingham 提出了隐私控制器架构[4]。使用语义 Web 技术（如 XML、RDF 和本体）表示的数据被推理引擎放大。这些引擎可以执行规则处理或利用基于本体的推理从现有的数据（显式的）得出新的数据（隐含的）。在该架构中，使用语义 Web 技术描述隐私约束（即隐私策略）。如果隐私控制器发现推理出的新数据违背隐私策略，它将给用户提出建议，告诉用户哪些信息应该保持私有的。

针对由语义关联引发的隐私泄露问题，Gowadia 等提出了一种面向 XML 文档的、具有灵活安全控制粒度的访问控制框架[38]。在该框架中，使用 RDF 陈述表示安全对象并表达安全策略(即访问控制规则)。访问控制规则中的三元组（s, o, ±a）对应于 RDF 陈述中的主体、客体和访问类型属性。RDF 陈述增强了规则的语义，增加了模型的灵活性和描述能力。特别是 RDF 提供了一种可推理的方式以表达基于关联的约束，防止语义关联隐私泄露问题。

6.3 语义 Web 信任管理

语义 Web 层次架构中的最高层是信任层，也就是说，如何信任语义 Web 提供的信息。Thuraisingham 详细分析了语义 Web 面临的信任问题[4]，例如，是否可以信任语义 Web 上的信息以及数据源，对包含不可信任数据的推理结果是否值得信任，如何协调各方（比如 Agent）之间的信任并达成合约，如何将信任管理和协商的构造融入语义 Web 技术中，信任管理的语义是什么等。Richardson 分析了利用语义 Web 技术的 Agent 信任管理问题[39]。

信任管理的一个重要方面就是信任协商。双方可以彼此协商信任值以及在它们之间共享的数据。目前有 2 种信任协商机制：基于凭据的信任协商和基于声誉的信任协商。在基于凭据的信任管理方面，Bertino 和她的团队已经进行了广泛的研究[40]。其思想是，交换各方的凭据，并依赖于凭据的类型，在双方之间建立信任。初始情况下，通过一些凭据机构获得凭据。因此，如果 John 想要看 Jane 的个人数据，他必须向 Jane 出示某凭据机构提供给他的凭据。在基于声誉的系统中，基于某人过去的行为所获得的声誉分配信任。例如，如果 Jane 申请教师职位，那么那些听说过 Jane 的人将讨论她的声誉，诸如她是不可靠的，她大量的时间缺课。如果这是事实，那么 Jane 作为教师的声誉是不好的，因此 Jane 不会被信任去给这份工作。在日常生活中，大家一直使用声誉。也就是说，人们是基于声誉来信任一个人或者一个组织。提高声誉，通常是很难的。然而，毁坏声誉却很容易，并会因此而降低信任值。在文献[41]中讨论了基于声誉的信任系统。

普渡大学的 Bertino 和她的小组开发了一个称为 Trust-X 的对等（peer-to-peer）框架用于信任的建立，它是一个基于凭据的信任系统。Trust-X 使用 XML 用于信任策略的表示[42]。

6.4 语义 Web 策略规范

策略可以指导或影响某个领域中实体的行为方式，策略为实体的行为提供规则。通过将策略与机制分离，可以动态地改变实体的行为而不改变实现。在信息安全领域，根据需求和安全模型，常用策略通常有访问控制策略、隐私保护策略、信任策略、完整性策略、会话策略、行为策略、管理策略等各种类型的策略。虽然语义 Web 是现有 Web 的一种演化，但语义 Web 具有其特有的技术组件和实现架构，用于现有 Web 的策略规范并不完全适合语义 Web 环境。为此，需要探索可应用于语义 Web 场景的新策略表示语言和规范。

普渡大学的 Bertino 和她的小组开发的 Trust-X 对等框架使用 XML 用于信任策略的表示[42]。W3C 推荐的 P3P 策略也使用具有命名空间的 XML 来编码策略。尽管 XML 是一种适合的策略语言，但它有一些缺陷：不能充分地表示语义。例如，语句“仅仅 B 不信任 C，A 才信任 B”或者“A 信任 B，而

且 B 信任 C，不意味着 A 信任 C”，用 XML 很难表达这样的陈述。同样，如果想要表达语义更丰富的隐私策略或者基于策略的推理能力，XML 的能力也是有限的[4]。因此需要表达能力更丰富的语言来描述信任和隐私策略。

在策略方面的大多数研究局限于某个应用领域，例如，安全、网络管理等，而没有通用的策略规范。策略的另一个问题是它们需要领域相关的某些信息，这就强制研究人员创建绑定于某个领域的策略语言。这阻碍了策略语言的灵活性和跨领域的可用性。为此，Kagal 等基于道义逻辑（deontic logic）建模了一种名为 Rei 的通用策略语言规范[27,28]，并开发了对应的策略引擎。Rei 的发音为 ray，它是一个日语单词，在日语中意味着通用的或本质，表明该策略语言的普遍可用性和灵活性，可以跨领域描述各种各样的策略，比如安全策略、会话策略和行为策略。Rei 包含权限（rights）、禁止（prohibitions）、义务（obligations）和豁免（dispensations）的概念。该策略语言使用一阶谓词逻辑（first order logic）进行描述，可以容易地与 RDF、DAML+OIL 和 OWL 之间进行互相转换。

语义 Web 语言（如 OWL/本体、SWRL 等）本身具有很强的描述以及推理能力。尤其 SWRL 在指定策略方面越来越流行，SWRL 集成了 RuleML 和 OWL，并具有其两者的优势。本质上，SWRL 具有 OWL 的表示能力以及 RuleML 的推理能力。Thuraisingham 尝试了分别使用 XML、RDF、OWL/本体和类似 RuleML 的语言作为策略规范来表达安全策略、隐私策略、信任策略、完整性策略以及在数据世系（provenance）中标注数据的历史[4]。同时，Thuraisingham 正在研究如果将查询重写规则应用到 SPARQL 中或者直接使用 SPARQL 来表达安全策略，以确保 RDF 数据库的访问安全。

SWRL 的提出是在 OWL 2 出现之前为了增强 OWL 1.0 的描述能力，并成为语义 Web 规则事实上的标准语言。虽然后来 OWL 2 的出现弥补了 OWL 1 在表达能力上的不足，但仍然缺乏 SWRL 的一些表达和推理特性。一些研究者正在探索结合本体构造和 SWRL 来作为面向语义 Web 环境下的策略规范。在 Masoumzadeh 等提出的一种基于本体的访问控制模型中[32,33]，结合社会化网络系统本体、访问控制本体和 SWRL 来表达访问控制策略规则，基于语义 Web 推理工具（例如 Jena、Pellet 等）还可以推理策略的一致性。在 Javanmardi 等提出的用于保护语义 Web 资源的基于语义的访问控制模型[31]中，基于 OWL 本体语言描述主体之间、客体之间以及行为之间的语义关系；为了增强授权规则的表达能力，SBAC 采用了 SWRL 规则语言。

在语义 Web 环境下，基于语义丰富的本体来描述资源的元数据，这些资源包括访问主体、被访问对象，它们都位于某个语义上下文中。因此，在表达访问控制需求时，仅仅通过简单的数据标识或者谓词来表示访问主体和访问对象是不够的，而应该根据语义元数据来描述它们。Damiani 等通过基于本体的语义元数据来扩展当前的标准策略语言 SAML 和 XACML 以表达访问控制规则，并提出了一个用于执行语义感知的策略的参考架构[43]。

6.5　语义 Web 数据存储安全

Oracle 数据库通过 VPD 和 OLS 技术提供了对存储语义数据的安全管理。Thuraisingham 提出了基于语义 Web 语言描述的安全策略对存储语义数据的数据库访问操作进行透明地修改的思想[4]。例如，根据安全策略，检查用户授权。假定用户要访问的内容包含未授权的部分，将根据安全策略，透明地重写查询语句，屏蔽用户未被授权访问的内容。同时，Thuraisingham 正在研究如果将查询重写规则应用到 SPARQL 中或者直接使用 SPARQL 来表达安全策略，以确保语义数据库的访问安全。

6.6　语义 Web 服务安全

基于语义 Web 标准构建的语义 Web services，使用语义 Web 语言表示在消息及 Web 服务描述语言中的语义。利用本体处理异质性，例如，如果在消息或服务描述中的词汇理解起来模棱两可，那么本体可以解决这样的歧义。使用规则语言（例如 SWRL）增强消息和服务描述的推理能力，使得编程人员集合来自不同源和服务的数据变得更加容易，而且不会丢失语义。语义 Web services 提供了机器可理解的 Web services，为实现自动化地 Web services 发现、调用和组合提供了有效的解决方案。但语义 Web services 面临与传统 Web services 同样的安全、隐私和信任问题。

Thuraisingham 分析了语义 Web services 的安全问题[4]。安全的语义 Web services，首先需要利用安全的语义 Web 技术，即语义 Web services 使用的语义 Web 技术组件必须是安全的。其次，可以利用语义 Web 语言的描述和推理能力来表示语义 Web

services 的安全、隐私和信任需求或策略。最后，可以将语义 Web services 的安全、隐私和信任管理实现为语义 Web services。

6.7 语义 Web 安全标准

为使语义 Web 技术获得广泛应用并创造真正的价值，除了解决语义 Web 技术当前面临的技术难题以外，一个不容忽视而且首先需要解决的问题是语义 Web 的安全问题。语义 Web 安全问题的解决不仅仅是一个技术层面的问题，还需要依赖相应的安全标准规范的建立。目前还没有专门用于语义 Web 环境的安全标准，这需要政府组织、学术界、标准化组织、IT 产业界一起共同努力，逐步建立并完善语义 Web 的相关安全标准。

7 结束语

语义 Web 技术虽然在技术和标准化方面都取得了一定的进展，但还面临着诸多技术挑战，距离语义 Web 远景的实现仍然还有一段距离。尤其是语义 Web 的安全问题必须首先解决，才能推动语义 Web 技术的广泛应用和发展。语义 Web 安全问题的解决，一方面需要结合传统 Web 安全解决方案和语义 Web 的特性探索适用于语义 Web 新的解决方案，另一方面，语义 Web 的安全问题不仅仅是技术问题，还涉及到安全规范、法律法规约束以及政府的监管，需要政府组织、学术界、标准化组织以及 IT 产业界共同努力才能最终解决。

参考文献：

[1] ANTONIOU G, HARMELEN F V. A Semantic Web Primer[M]. Cambridge: MIT Press, 2008.

[2] BERNERS-LEE T, HENDLER J, LASSILA O. The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities[J]. Scientific American, 2001, 284(5): 34-43.

[3] Foaf project[EB/OL]. http://www.foaf-project.org, 2010.

[4] THURAISINGHAM B. Secure Semantic Service-Oriented Systems[M]. Boca Raton, FL: CRC Press, 2011.

[5] Semantic web-XML2000[EB/OL]. http://www.w3.org/2000/Talks/ 1206-xml2k-tbl/Overview.html, 2000.

[6] RDF Primer[S]. http://www.w3.org/TR/rdf-primer/, 2004.

[7] RDF Vocabulary Description Language 1.0: RDF Schema[S]. http:// www.w3.org/TR/rdf-schema/, 2004.

[8] OWL 2 Web ontology language primer[EB/OL]. http://www.w3.org/ TR/ owl2-primer/, 2009.

[9] FENSEL D. The semantic Web and its languages[J]. IEEE Intelligent Systems, 2000, 15(6):67-63.

[10] DAML+oil[EB/OL]. http://www.daml.org/2001/03/ daml+ oil-index.html, 2001.

[11] DAML-ONT[EB/OL]. http://www.daml.org/2000/10/daml-ont.html, 2000.

[12] OIL[EB/OL]. http://www.cs.man.ac.uk/~horrocks/OIL/Semantics/, 2000.

[13] SWRL: A semantic Web rule language combining OWL and rule ML[EB/OL]. http://www.w3.org/Submission/SWRL/, 2004.

[14] HEBELER J, FISHER M, BLACE R, *et al*. Semantic Web Programming[M]. Indianapolis: Wiley Publishing, 2009.

[15] SPARQL query language for RDF[EB/OL]. http://www.w3.org/TR/rdf-sparql-query/, 2008.

[16] XML encryption syntax and processing[EB/OL]. http://www.w3.org/ TR/xmlenc-core/, 2002.

[17] XML key management specification (XKMS)[EB/OL]. http://www. w3.org/ TR/xkms/, 2001.

[18] XML-signature syntax and processing[EB/OL]. http://www.w3.org/ TR/ xmldsig- core/, 2002.

[19] Assertions and protocols for the OASIS security assertion markup language(SAML) V2.0[EB/OL]. http://docs.oasis-open.org/security/saml/v2.0/, 2005.

[20] EXtensible access control markup language(XACML) version 2.0[EB/OL]. http://docs.oasis-open.org/xacml/2.0/, 2004.

[21] OWL Web ontology language guide[EB/OL]. http://www.w3.org/TR/owl- guide/, 2004.

[22] DARPA agent markup language (DAML) program[EB/OL]. http:// www. daml.org/, 2006.

[23] The EU's sixth framework programme sixth framework programme for research and technological development[EB/OL]. http://ec.europa.eu/research/fp6/index_en.cfm, 2006.

[24] The EU's sixth framework programme seventh framework programme for research and technological development[EB/OL]. http://cordis.europa.eu/fp7/home_en.html, 2012.

[25] PM donates tim berners-lee's body to web science[EB/OL]. http://www.techradar.com/news/internet/gordon-brown-donates-tim-berners-lee-s-life-to-web-science-678658?src=rss&attr=all, 2010.

[26] P3P project[EB/OL]. http://www. w3.org/P3P/, 2006.

[27] KAGAL L, FININ T, JOSHI A. A policy language for a pervasive computing environment[A]. Proceeding of 4th IEEE International Workshop on Policies for Distributed Systems and Networks[C]. Lake Como, Italy, 2003. 63-74.

[28] KAGAL L. Rei: A policy Language for the Me-centric Project[R]. 2002.

[29] BERTINO E, CASTANO S, FERRARI E, *et al*. Protection and administration of XML data source[J]. Data and Knowledge Engineering, 2002, 43(3):237-260.

[30] BERTINO E, CARMINATI B, FERRARI E, *et al*. Selective and

authentic third party publication of XML documents[J]. IEEE Transactions on Knowledge and Data Engineering, 2004, 16(10):1263-1278.

[31] JAVANMARDI S, AMINI M, JALILI R, *et al*. SBAC: A semantic based access control model[A]. Proceeding of the 11th Nordic Workshop on Secure IT-systems (NordSec2006)[C]. Linkping, Sweden, 2006. 157-168.

[32] MASOUMZADEH A, JOSHI J. Ontology-based access control for social network systems[J]. International Journal of Information Privacy, Security and Integrity, 2011, 1(1): 59-78.

[33] MASOUMZADEH A, JOSHI J. OSNAC: an ontology-based access control model for social networking systems[A]. Social Computing (SocialCom) 2010 IEEE Second International Conference[C]. Minneapolis, MN, USA, 2010. 751-759.

[34] CAMPANA F, ANNICCHIARICO R, RIAÑO D, *et al*. The K4CARE Model[R]. 2007.

[35] GIBERT K, VALLS A, CASALS J, *et al*. Sample APOs[R]. 2007.

[36] CARMINATI B, FERRARI E, THURAISINGHAM B. Using RDF for policy specification and enforcement[A]. Proceedings of the DEXA Conference Workshop on Web Semantics[C]. Zaragoza, Spain, 2004. 163-167.

[37] NIVARGI P. A Generic Privacy Model for Data Access Using Semantic Web Technologies[D]. Metro Phoenix, Arizona; Arizona State University, 2004.

[38] GOWADIA V, FARKAS C. RDF metadata for XML access control[A]. Proceedings of the 2003 ACM Workshop on XML Security[C]. New York, NY, USA, 2003. 39-48.

[39] RICHARDSON M, AGRAWAL R, DOMINGOS P. Trust management for the semantic Web[A]. Proceeding of the Second International Semantic Web Conference[C]. Sanibel Island, FL, 2003. 351-368.

[40] BERTINO E. Trust-X: an XML framework for trust negotiations[A]. 7th IFIP TC-6 Conference on Communications and Multimedia Security[C]. Torino, Italy, 2003. 146-157.

[41] SHMATIKOV V, TALCOTT C. Reputation-based trust management[J]. Journal of Computer Security, 2005, 13(1):167-190.

[42] BERTINO E, FERRARI E, SQUICCIARINI A C, *et al*. Trust-X: a peer-to-peer framework for trust establishment[J]. IEEE Transactions on Knowledge and Data Engineering, 2004, 16(7):827-842.

[43] DAMIANI E, DI VIMERCATI SDC, FUGAZZA C, *et al*. Extending policy languages to the semantic Web[A]. International conference on web engineering (ICWE2004)[C]. Munich, Germany, 2004. 330-343

作者简介：

陈德彦（1977-），男，四川南充人，东北大学博士生，主要研究方向为网络与信息安全、语义 Web、云计算等。

赵宏（1954-），男，河北河间人，博士，东北大学教授、博士生导师，计算机软件国家工程研究中心副主任，主要研究方向为下一代网络、网络与信息安全和网络管理。

张霞（1965-），女，辽宁沈阳人，博士，东北大学教授，东软集团股份有限公司软件架构国家重点实验室主任，主要研究方向为软件架构、软件工程、数据库技术等。

赵立军（1971-），男，辽宁沈阳人，硕士，东软集团股份有限公司工程师，主要研究方向为软件架构。

平安（1965-），男，河北定州人，博士，东软集团股份有限公司软件架构国家重点实验室副研究员，主要研究方向为软件架构。