

Internet 应用的普及和软件技术的进步使基于 Web 的网络计算模式得到了广泛的重视和应用, 面向 Web 应用的三层结构 (或多层结构)、构件化的商务逻辑封装、基于 XML 和 Web Services 的自动电子商务等理论和技术已深入人心并广泛应用。但创新是无止尽的! 实际上, 现有的这些 Web 技术并不能真正实现人们期待已久的智能搜索引擎、智能信息代理、智能交易代理等等基于 Web 的个性化、智能化的应用。而这, 正是语义 Web (Semantic Web) 要解决的问题。

## 智能化应用的基石

——基于本体的语义 Web 知识处理平台

清华大学计算机系

徐剑军 张钊 李涓子 王克宏

### 语义 Web

语义 Web 就是建立一个以本体为基础的具有语义特征的网络, 使计算机能够真正理解数据的语义。随着语义 Web 在实际应用中的推广, 人们期待已久的各种智能化网络应用将成为现实。表 1 总结了 Web 技术的发展和语义 Web 在其中的位置。

表 1 Web 技术的发展

	第一阶段	第二阶段	第三阶段
--	------	------	------

主要的信息描述格式	HTML	XML 和在此基础上建立的领域标准化格式，如 ebXML 等。	RDF DAML
信息描述特点	非格式化或半格式化。	信息格式化、标准化。信息定义缺乏关联，有可能有多重含义。	基于本体的信息定义与表达，实现语义 Web。
动态交互模式与技术	处理人机交互的 CGI 技术和类似的动态脚本语言 CGI、ASP、JSP 等。	Web 服务、SOAP、UDDI、WSDL。	语义 Web 服务具体模式和技术有待发展。
动态交互特点	处理人——机基于 Internet 的交互，但计算机与计算机之间难以交互。	可处理计算机——计算机的交互，交互过程自动化，但难以构建个性化、智能化的服务。	协调处理人——机交互与计算机对计算机的交互，交互过程智能化，可提供智能化、个性化的服务。
发展现状	应用广泛	技术已日益成熟，已有大量相关技术规范，行业应用正在推广中。	从具体模式、相应理论到实现技术、行业应用都有待发展。

“本体论”最早是哲学中的基本概念，它是研究“是”之所以为“是”的理论，可以说是哲学中的哲学，甚至可以认为西方哲学的发展就是一个“本体论”的产生、发展、怀疑和批判的过程。如果将本体技术引入 Web 技术中，由于本体具有丰富的语义和广泛的关系，那么将从根本上解决目前 Web 的信息格式的异构性、信息语义的多重性以及信息关系的匮乏和非统一性。它将变革现有的 Web 服务，使之成为智能化的语义 Web 服务，使 Web 实现从自动化到智能化的转变成为可能。

### 语言标准与平台系统

基于本体的语义表示是语义 Web 中一个核心任务。XML 是语义表示的基础。XML 句法层通过使用内嵌和可带有属性的元素表示信息，但这些标记信息不具有解释数据语义内容的能力。这时由 W3C 制定的 RDF 就成为 XML 之后的 Web 信息表示基础标准。

RDF (Resource Description Framework) 是一个建立在 XML 技术之上的信

息（知识）描述标准。RDF 数据层用于描述资源、链接及其类型，允许对结构元数据进行译码、交换和重用。它的文档是一个由主谓宾结构的断言的集合且包括一个数据模型 (RDFMS)和一个 schema (RDFS)。

DAML (DARPA Agent Markup Language) 发布于 2000 年 8 月，它的目的是为本体层提供语言和工具。它属于军方 DARPA (Internet 就是源自该组织的研究项目) 的计划。DAML 是在 W3C 的 XML 与 RDF 标准基础上建立的。DAML 是通过自己表达本体，同时对原有 Web 页面进行注解，链接到相关的本体上的。

OWL (Web Ontology Language) 是 W3C 网络本体工作组设计一种的网络本体语言。它包含一个高层的抽象语法，它充分利用了 DAML，可以看成是 DAML 的一个升级。并且一个有理论模型的语义形式化表示。可以对 OWL 本体进行形式化表示，并且可以映射到 RDF 的语法。

以上制定的标准虽然从形式上描述了语义 Web 上的本体，但是要具体实现各种基于语义 Web 应用还有一定的距离，还需要各种语义 Web 的知识处理平台以达到这个目标。所谓平台，其自身并不一定是一个完整的应用程序，而是可以通过向行业应用系统提供方便易用的基本处理接口，从而可以处理本体形式化表示、存储管理、分布式管理、甚至是推理机制等各种基础功能。下图说明了语义 Web 的结构



图 1 语义 Web 的总体结构

### 知识处理平台

语义 Web 的实现需要支撑软件的支持，因此很多科研机构、学校、公司在语义 Web、DAML 及 RDF 基础上开展了研究工作，开发了许多工具和平台。

Sesame 是一个开放源代码的本体存储与查询系统，由 AIdministrator 公司开发。它解决了以往的本体数据存储的问题，支持关系型数据库和面向对象的数据库。并且实现了 RQL (RDF Query Language) 查询语言，可以在数据库中检索需要的本体。

KAON (The Karlsruhe Ontology and Semantic Web Infrastructure) 是德国 Karlsruhe 大学的一个科研项目，致力于语义 Web 提供所需的基础本体系统和相关工具。它针对基于本体的上层商业应用的需求提供了一个开放的本体管理基础

软件，为本体的存储、创建、标识提供了一个全面的工具集。

HP 实验室的 Jena 是一个以 RDF API 为核心的语义 Web 工具。Jena 由 Network API、Query、Readers、Inference、Writers、Stores 几部分模块围绕 RDF API 组成。它支持 RDF 的创建、操作和查询等功能。而 RDF API 也能够支持很多不同的数据存储技术。插件式的接口设计能够自动适应各种用不同的语言编写的读写 RDF 文件的需要。推理层、查询功能以及网络 API 建立在 RDF API 之上。

WODOS (Web Oriented Distributed Ontology System) 系统由清华大学知识工程室研究开发。它是一个全面的基于本体的语义 Web 知识处理平台，支持中英文，已经可将本体信息存于关系型数据库中，可以做 RDF 文件的导入导出，支持 RQL 查询，第三方软件可以通过 OdoAPI 进行操作。

几乎从 Internet 开始普及推广的时候起，人们就一直在描绘智能搜索引擎、个性化智能信息推荐、智能代理、智能电子商务、个性化电子服务、基于 Web 的 KOD (Knowledge on Command) 等的美好前景。但这些从来就没有应用推广过，它们就像“科幻小说”一般遥不可及，其主要原因是 Web 智能服务所必需的面向 Web 的基于本体的信息及知识表达相关标准、理论和技术还有待发展。

另一方面，本体论在人工智能中的运用也由来已久，但在实际应用中并没有广泛的效果。Web 的出现为本体论提出了新的、更广泛的需求，也提供了一个建立真正全球化的、统一的、标准化的本体环境的途径。

#### **相关链接**    **WODOS 的实现**

WODOS 系统的功能模块规划成四个部分，这四个部分相辅相成，各自都有明确的功能。

首先，系统要提供本体信息的数据库存储机制，这就需要本体数据库的访问操作模块，我们定义为 RCS；其次，作为支撑软件，WODOS 是以第三方软件的形式提供给别人使用的，这就需要定义一个对外的接口，这个接口包括一个 OdoAPI 本地本体访问模块和一个 OdoAPI 查询模块（含 RQL 解释器），整个接口模块我们定义为 ACS；第三，处理本体需要很多的参数，而且要调用系统的功能需要一个可视化的界面，我们将这一部分界面和管理功能模块定义为 MCS；最后，为了实现基于本体的推理机制，我们定义了以 F-Logic 推理内核为核心的功能模块 ICS。WODOS 整体结构如图所示。

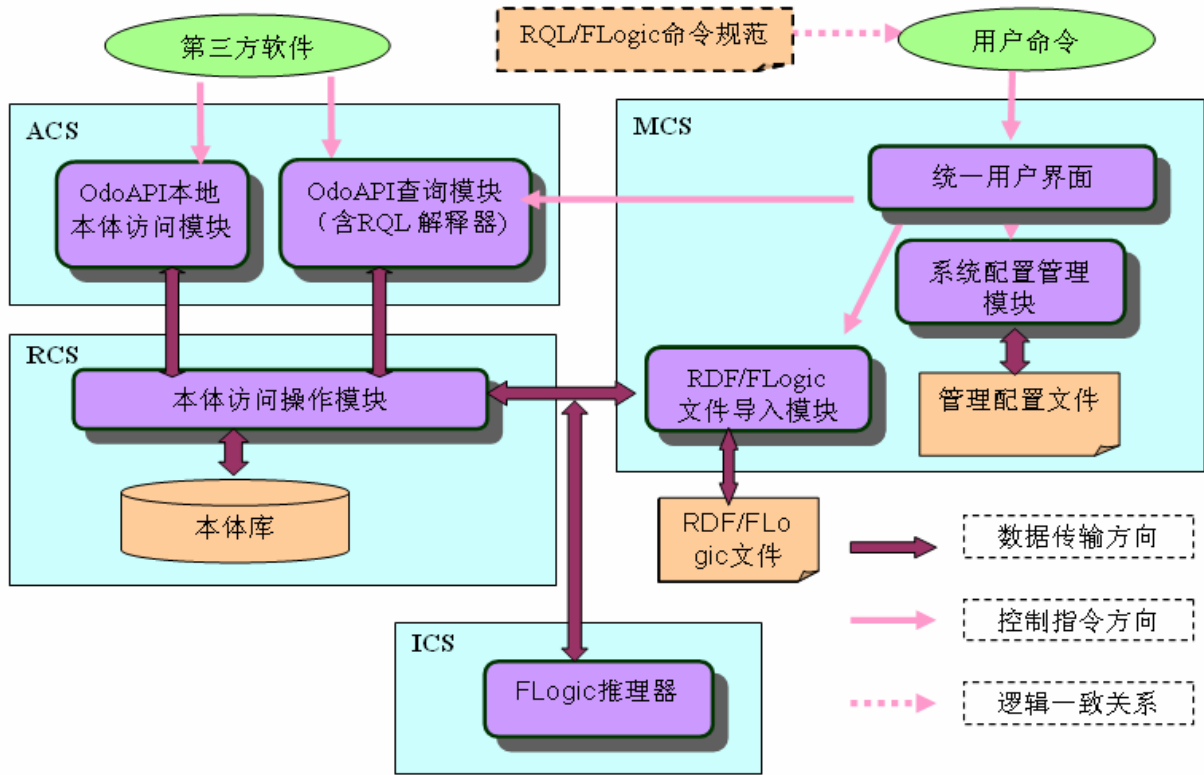


图 WODOS 整体结构图