

国外语义网发展概述

张 剑

中国科学院文献情报中心 北京 100080

中国科学院研究生院 北京 100039

〔摘要〕通过查阅相关文献及调研 WWW、ISWC 有关语义网的国际会议,从相关标准、语义网工具、研究项目三个方面分别概述国外自 2001 年以来有关语义网的研究进展,揭示并简要分析目前语义网研究存在的问题及所面临的挑战,为国内语义网的开发提供一些借鉴。

〔关键词〕语义网 W3C 语义网研究

〔分类号〕TP18

New Development of Overseas Semantic Web

Zhang Jian

Library of the Chinese Academy of Sciences, Beijing 100080

Graduate School of the Chinese Academy of Sciences, Beijing 100039

〔Abstract〕In recent years, the study of semantic web has become a research focus at home and abroad. By analyzing the international conferences of WWW and ISWC, new progress on semantic web has been introduced from relative standards, semantic tools and related projects. In this paper, problems and challenges that the current semantic web study is facing are presented in order to show prospectives of the development of semantic web.

〔Keywords〕semantic web W3C research of semantic web

语义网 (semantic Web) 自 WWW 的创始人提姆·伯纳斯李 (Tim Berners-Lee) 在 2001 年正式提出后,立即引起了国内外学者的极大兴趣。国内许多研究机构和大学,如中国科学院软件研究所电子商务技术研究中心、国防科技大学计算机系、浙江大学、上海交通大学、北京大学等都先后对语义网进行了相关研究,W3C (万维网联盟) 是专门致力于创建 Web 相关技术标准并促进 Web 向更深、更广发展的国际组织,该组织一直以引领 Web 技术的发展和促进为己任,其成员已经达到 400 个机构和团体。国际语义网研讨会 (International Semantic Web Conference, 以下简称 ISWC) 自 2001 年以来每年召开一次会议,其内容覆盖了语义网的方方面面,因此,跟踪这两个权威性的国际会议,足以反映语义网的最新发展。

1 相关标准

2004 年 2 月 10 日,W3C 规定 RDF 和 OWL 作为 W3C 推荐级标准,RDF 表示信息和网络中的知识交换,而 OWL 则表示本体的发布和共享,它同时支持高级的网络检索、软件代理和知识管理。相关标准如下:

- 资源描述框架/可扩展标记语言的语法规则 (修订

版) (RDF/XML Syntax Specification (Revised)): <http://www.w3.org/TR/rdf-syntax-grammar/>

- 资源描述框架词汇描述语言 1.0: 资源描述框架结构 (RDF Vocabulary Description Language 1.0: RDF Schema): <http://www.w3.org/TR/rdf-schema/>

- 资源描述框架概要 (RDF Primer): <http://www.w3.org/TR/rdf-primer/>

- 资源描述框架 (RDF): 概念与定义方式 (Resource Description Framework (RDF): Concepts and Abstract Syntax): <http://www.w3.org/TR/rdf-concepts/>

- 资源描述框架语义 (RDF Semantics): <http://www.w3.org/TR/rdf-ml/>

- 资源描述框架测试案例 (RDF Test Cases): <http://www.w3.org/TR/rdf-testcases/>

- 网络实用分类系统语言使用案例与要求 (Web Ontology Language (OWL) Use Cases and Requirements): <http://www.w3.org/TR/rdf-testcases/>

- 网络实用分类系统语言参照 (OWL Reference): <http://www.w3.org/TR/owl-ref/>

- 网络实用分类系统语言语义及定义方式 (OWL Se-

manatics and Abstract Syntax); <http://www.w3.org/TR/owl-ref/> www.w3.org/TR/owl-features/

- 网络实用分类系统语言综述 (OWL Overview); <http://www.w3.org/TR/owl-features/>
- 网络实用分类系统语言测试案例 (OWL Test Cases); <http://www.w3.org/TR/owl-features/>
- 网络实用分类系统语言指南 (OWL Guide); <http://www.w3.org/TR/owl-features/>

2 语义网工具

随着语义网研究的不断深入和发展,出现了许多新的工具,本文对各种工具进行了简单的归类,详细情况请参见表 1。

表 1 语义网工具列表

类型	名称	网址	简介
图形编辑器	IsaViz (RDF 的可视图形创作工具)	http://www.w3.org/2001/11/IsaViz/	一个 2.5D 的用户界面,可以在图片中自由滑动和检测通过画椭圆、方框、弧线等来创建和编辑图形 RDF/XML, Notation 3 and N-Triple 输入 RDF/XML, Notation 3 and N-Triple 输出, 也有 SVG 和 PNC 输出 自第二版后, IsaViz 能够使用 CSS (表示从 CSS 和 SVG 演化来的结点连接图) 编辑 RDF 图形
RDF 环境 (翻译器等)	Jena	http://jena.sourceforge.net/	是一个 java 开发工具包,用于开发语义网的应用系统 由 HP 公司的 Brian McBride 开发,起源于 SIRPAC API 其本体解析器包括三部分,其中包括对 RDQL 的查询支持和对 OWL 的解析。对 OWL 而言,语义逻辑的处理才是推理机智的实现,最新版本的 Jena 已支持 OWL 的语义逻辑的处理 Jena2.1 提供的对 OWL 的支持包括:方便地访问标准 OWL 的类及其属性;支持多种版本的 OWL 规范在基本的查询中通过 subclassOf 来实现类的层级访问和使用;可以注册用来映射 xmlschema 的数据类型和 java 对象的转换器;支持基本的对 list 的处理;自动处理本体中 imports 的 statement;识别传递属性和互斥属性
	Redland (Redland 是一套免费的 RDF 软件包)	http://www.redlandopensource.ac.uk/	是用 C 语言编写的 能够处理 RDF 图像、三元组 (triples)、URLs 和 Literals 的应用程序接口 用 Sleepycat/Berkeley DB、MySQL 3/4、AKT Triplestore_files 或 URLs 来存储图片 其剖析器和串行器 (把并行数据转换成串行数据) 通过 RDF 剖析工具包把 RDF 文档转换为 RDF/XML、N-Triples 和 Turtle Terse RDF 三元组语句句法格式来进行读写
	Kowari	http://www.kowari.org/	是用 Java 1.4 编写的 作为一个 3 倍 (triples) 数据库而建立起来的 支持 Jena API、JRDF、可选择性的 RDF 对 API (为这些 API 增加新的类 SQL 查询语言)、iTQL 等进行管理
	Sesame	http://sourceforge.net/projects/sesame/	是 RDF 存储、提问和推论的一个基础模型 能够展开为一个 Web 服务或被用作一个 Java 库 包含 3 种提问语言 (SeROL, RQL, RDQL)、支持推理和连续性存储
合格检测器	针对 RDF 的合格检测器	http://www.w3.org/RDF/Validator/	这种在线服务已被更新,目前它支持 RDF 核心工作组发布的《规格说明书》的最终工作草案,包括数据类型 不支持《RDF 模型与句法规格说明书》中的反义元素和属性 只需在页面上的文本框中键入一个 URL 或粘贴一个 RDF/XML 文档,即可将显示相应的数据模型用一个三元组或者以图的形式来表示
	针对 OWL 的合格检测器	http://owl.bbn.com/validator/	一种用来检查 OWL 中简单的句法错误的工具。能阅读 OWL 文件并且检查各种潜在错误,同时能提供错误及警告清单、标注它们在文章中的位置,并为错误更正提供一些指导性帮助。但所输入的文章必须是有效的 RDF 文档 不具有完整的推理机制,如文章中使用了较多 OWL 的高级特性,其检测结果可能不完全准确。有时它能提醒你无问题的路径或位置 目前该检测器与 OWL-DL 粗略地保持一致
本体论转换器	针对 OWL 的本体论置换器	http://www.mindswap.org/2002/owl.html	把 DAML + OWL 文件转换成 OWL 文件 输入你 DAML 文件的 URL,包括该文件新缺省的命名空间 有一个 OWL 文件输出,可以直接存储
其他软件	Protégé-2000	http://protege.stanford.edu/	由美国斯坦福大学医学院开发 用 Java 和 Open Source 作为操作平台 可用于编制实用分类系统和知识库 (knowledge base),有可自行设置的数据输入格式,并能输入数据,也可用插件来扩展一些如提问、XML 转换等特殊的功能 输出格式有文本、HTML、JDBC、RDF Schema 及 XML Schema
	OntoEdit	http://www.ontoprise.de/	德国的 Ontoprise 公司开发 提供实用分类系统工程环境,支持构造概念、关系、定理,不依赖于某一语言表述 利用模块和插件结构,灵活性强,能够方便地引进专门功能和词库 支持 RDF 和 DAML 国防高级研究项目部智能代理置标语言 (DARPA Agent Markup Language),并能输入和输出数据库结构与数据 (如 Oracle, MSSQL, DB2)
	OilEd	http://oiled.man.ac.uk/index.shtml	由英国曼彻斯特大学开发 支持表述语言 DAML + OIL (Ontology Inference Layer, 实用分类系统交换层)。该软件的主要目的是提供一个“笔记簿”式的编辑软件来示范术语快速分类 (Fast Classification of Terminology, 简称 FaCT) 的推理算法
	RDF Gateway	http://www.intellidimension.com/	该软件与应用服务器、网络服务器以及归纳性 RDF 数据库服务器一道,提供类似程序设计模型与语言功能,使网络与数据库开发能够迅速地开始语义网的应用

3 研究项目

3.1 语义网技术在美国的应用

在美国医学界,语义网技术已经被应用于描述医学术语,针对各种编码方案从不同角度和目标用语义网语言对复杂的医学文档进行描述。作为 MetaheSAWs 计划的一部分,美国国立癌症研究所的生物信息学中心正在将一个庞大的癌症研究专业词汇表转化为一个可机读的“概念集”,其本质就是一个扩展了的辞典,它能精确描述词汇表中的每一个词条与以资源定义框架为基础建立的网络概念集语言(OWL)间的关系。它还可以提供一个癌基因的扩展定义,定位癌基因所在的生物体、生物体功能及相关疾病。更为重要的是,其他如 PubMed 等在线资源也可以同这类“概念集”中所收录的术语建立链接。

美国国立虚拟天文台正在探索用语义网技术将数量众多的天文学资源全部链接起来。另外,经过万维网协会和其他一些电脑组织的大力推广,语义网已初见成效。美国国防部高级研究计划局(the Defense Advanced Research Projects Agency,简称 DARPA)为语义网项目先后捐助了 3 500 万美元,并为美国国防部开发了叫做 DARPA 的代理标记语言(DARPA Agent Markup Language),这种基于语义网的语言允许用户在 Web 文件中加入其他类型数据,并把它们链接到信息实体上。

惠普公司投资的 SWARA 项目主要研究语义网与多媒体再利用技术是如何丰富网络信息资源的使用和服务的。

3.2 语义网在欧洲的发展

Ontoweb 是一个由欧洲委员会投资创建的主题网络,它的主要目标是把基于 ontology 的方法和语义网工具等方面结合在一起,是不同兴趣小组相互交流的平台。

SWAD - Europe 项目起源于 W3C 的语义网,项目起止时间是从 2002 年 5 月到 2004 年 10 月,主要目的是通过实施、研究和测试等活动来促进 W3C 语义网的发展,使 W3C 的语义网横跨欧洲各地。它把原来麻省理工学院(MIT)SWAD 工作所采用的方法推广到语义网兴趣小组的协同工作中,为语义网的发展建立了一个更广泛的基础。兴趣小组提供一个论坛,其主要目的是把 SWAD - Europe(包括欧洲或者其他地方与语义网相关的项目)和 W3C 的正式标准化活动联系起来。

德国国家人工智能研究中心(DFKI)的竞争力语义网(Competence Center Semantic Web,简称为 CCSW)提出要开发出一种能够从网络中半自动抽取语义元数据的方法,同时,该语义网还重视语义信息的集成,其目标是通过 DFKI 内各种语义网项目之间的合作达到一个协同效应。

3.3 语义网在其他方面的应用

随着语义网基础结构工具的出现,国外各大商业公司、科研机构也纷纷参与到语义网相关的研究当中。如诺基亚等公司也加入了语义网应用程序的开发;亚历山大大学研发的 KAON 为语义网提供了一个通用的应用框架。

另外,在实际应用方面,语义网也被广泛地应用到科研工作中,为科研工作中所产生的大量数据,如空间卫星扫描地面积累的数据、临床药物试验的数据、地质勘探数据和环境污染监测数据等提供编制元数据、检索、存档等服务。

在 WWW2004 会议上,Berners - Lee 对大会报告进行了总结。通过分析大会论文可知,语义网目前尚未完全成熟,但 OWL 和 RDF 已被限定为 W3C 的官方推荐标准,因而,语义网已有了坚实的基础,成功地完成了其第一个阶段的任务目标。目前,对语义网的研究正在进入“第二个阶段”,但对语义网的进一步发展尚未有一个完全清晰的前进路线图。语义网的一些框架原型(从 Jena 到 Haystack)虽能满足在商业上的应用,但只有个别的能面向终端用户,因此,语义网还面临着很多的问题和挑战。其主要问题是,随着系统规则的变化,计算机对互联网页面数据推论所需花费的时间会急剧增加。正如数学经典故事中所描述的:在确定一条穿过几个城市的最短路线时,推销人员很快即可找出所有可行路线中最佳的一条;但当城市数量变成 15 个时,可行路线将会增加 430 亿条以上。同样的问题也存在于利用计算机进行推断,强力搜索答案会导致时间上的浪费。目前,要解决的问题是如何才能使获得对储存于网络空间的数据进行智能评估的能力,为互联网上杂乱而庞大的信息设计一种构架,使它们易于查找和利用。

5 结 语

语义网在互联网上更多地是在搜索引擎中的应用,现有万维网是图像与文本的集合。语义学引入后将从根本上改变万维网的性质。语义网可将信息进行解释、交换和处理,并可从搜索服务器端通过分散于各处理服务器中的语义分析搜索数据库,从不同来源收集机器的可读数据中处理并推理出新的需搜索的数据。语义网可使分散于全球的成千上万的独立数据库融合,最终使用户独立运用 Internet 上庞大的信息资源。Tim Berners - Lee 在对未来的语义网进行展望时期望:语义网将成为一个能够理解人类语言的智能网络,它将可识别信息的意义,并自动对信息进行解释、交换和处理。比如,一台同语义网连接的计算机可将筛选出使用者所需的信息,并自动将这些信息传送到不同的装置和设备中去。

目前,语义网尚停留在研究阶段。国外虽然已经出现了第一个语义网络——Mindswap,但仍在着许多不足,值得国内外学者进一步探讨与试验。与国外相比,语义网在国内起

步
译
由
参
1
2
3
4
5
6
[4
—
(
12
13
14
[4

步较晚,所取得的成果也相对较少。但从总体来看,各国对语义网的研究基本上处于同一起跑线上,因而,我们应抓紧时机,再创网络辉煌。

参考文献:

- 1 Bernard-Lee T. WWW2004 Keynote. [2004-09-01]. <http://www.w3.org/2004/Talks/0519-tbl-keynote/Overview.html>
- 2 Bernard-Lee T. The Semantic Web—LCS seminar. [2004-09-01]. <http://www.w3c.org/2002/Talks/09-lcs-sweb-tbl/>
- 3 Curbera F, Khalaf R, Mukhi N. The next step in Web services. [2004-09-06]. <http://bbs.w3china.org/dragonstar/papers/Curbera+CACM46-10.pdf>
- 4 Lu S Y, Dong M, Fotouhi F. The semantic Web: Opportunities and challenges for next-generation web applications. [2004-09-08]. <http://InformationR.net/ir/7-4/paper134.html>
- 5 Ford P. WWW2004 Semantic Web Roundup. [2004-09-09]. <http://www.xml.com/pub/a/2004/05/26/www2004.html>
- 6 Dean M, Hendler J. Semantic web tools tutorial. [2004-09-12]. <http://www.daml.org/2003/05/swmu-tools-tutorial/Overview.html>

- 7 Semantic Web Working Symposium (SWWS). [2004-09-12]. <http://www.semanticweb.org/SWWS/>
- 8 1st International Semantic Web Conference (ISWC2002). [2004-09-14]. <http://iswc2002.semanticweb.org/>
- 9 2nd International Semantic Web Conference (ISWC2003). [2004-09-14]. <http://iswc2003.semanticweb.org/>
- 10 Semantic Web. [2004-09-16]. <http://www.semanticweb.org/>
- 11 SWAD - Europe. [2004-09-16]. <http://www.w3.org/2001/sw/Europe/>
- 12 Ontoweb. [2004-09-16]. <http://ontoweb.semanticweb.org/>
- 13 Competence Center Semantic Web. [2004-09-17]. <http://ccsw.dfki.de/>
- 14 秦健. 实用分类系统与语义网:发展现状和研究课题. 现代图书情报技术, 2004(1): 16-23
- 15 刘柏篱. 基于知识的语义网:概念、技术及挑战. 中国图书馆学报, 2003(2): 18-21
- 16 刘柏篱. 面向语义网的本体表示. 中国图书馆学报, 2004(2): 47-51

[作者简介] 张剑,女,1981年生,硕士研究生。

(上接第74页)

- 12 Choi S Y, Stahl D, Whinston B 著,张大力等译. 电子商务经济学. 北京:电子工业出版社, 2000: 1-144
- 13 Guen L H. AUCNET: Electronic intermediary for used-car transactions. Electronic Markets, 1997, 7(4): 24-28
- 14 Lee H G, Westland J C, Hong S. The impact of electronic marketplaces on product prices: an empirical study of AUCNET. International Journal of Electronic Commerce, 2000, 4(2): 45-60

- 15 Chiru A, Kauffman R. The "eBay of blank": digital intermediation in electronic commerce. In: E-Commerce and V-Business (eds. Barnes S J & Hunt B). London: Butterworth - Heinemann Publishers, 2000
- 16 Froomkin A M. The essential role of trusted third parties in electronic commerce. Oregon Law Review, 1996, 75(1): 49-116

[作者简介] 李莉,女,1970年生,博士,讲师,发表论文20余篇。
杨文胜,男,1969年生,博士,讲师,发表论文10余篇。
谢阳群,男,1962年生,教授,博士,发表论文50余篇。

下 期 要 目

- 从情报检索语言到本体——信息组织的新变革 (戴维民等)
- 学科信息门户信息组织的优化 (黄如花)
- 网络信息计量学软件及其开发方向探讨 (沙勇忠等)
- 1999-2004年国外图书馆学情报学研究领域的分析 (褚金涛)
- 基于ILM数字图书馆藏管理策略 (索传军)
- 资源环境学科信息门户的研究与建设 (祝忠明等)

- 上海城市的发展与图书馆发展的互动作用 (王世伟)
- 参考文献管理工具发展趋向略论 (沈熙)
- 试论网络实时参考咨询服务——兼述NSTL实时参考咨询系统的建设与服务 (顾德南)
- “公共知识中心”说辞难——与龚蛟腾等同志商榷 (张践明)
- 数字图书馆的资源整合 (龚亦农)