

语义 Web 服务的 OWL-S 描述及其应用

牟 帅, 黄映辉, 李冠宇

(大连海事大学 信息科学技术学院, 辽宁 大连 116026)

摘 要:实现语义 Web 服务的关键是对 Web 服务进行语义描述及其在语义检索模型中的应用。国外学术界提出的语义 Web 服务的 OWL-S 框架并没有充分考虑检索请求的中西语言表达的差异,因此该框架的实际应用必须对中文查询请求进行中文分词处理。语义 Web 服务的 OWL-S 检索模型,首先对中文检索请求语句进行中文分词,然后通过本体进行语义标注生成 Web 服务请求的 OWL-S 文档,继而在 OWL-S 语义扩展后的 UDDI 中实施语义检索匹配。该模型在中文分词的基础上,结合语义 Web 服务技术实现 Web 服务的动态查找与组合,可提高 Web 服务的查全率和查准率。

关键词:中文分词; Web 服务; OWL-S; UDDI; 语义检索

中图分类号:TP391.3

文献标识码:A

文章编号:1673-629X(2009)01-0013-04

OWL-S Description of Semantic Web Service and Its Applications

MU Shuai, HUANG Ying-hui, LI Guan-yu

(Information Science and Technology College, Dalian Maritime University, Dalian 116026, China)

Abstract:The semantic description of Web service and its applications in the semantic retrieval model are the key to realize the semantic Web service. The OWL-S framework of semantic Web service proposed by foreign scholars does not take the differences of the retrieval request expressed by Chinese or western languages into full consideration, so Chinese word segmentation should be done in actual applications. In the OWL-S retrieval model of semantic Web service, the Chinese retrieval request needs Chinese word segmenting firstly, semantic annotating though ontology secondly, then generating the OWL-S documents of Web service request, moreover, doing the semantic retrieval matching in the UDDI after semantic expansion by OWL-S. In the model, combining the semantic Web services technology on the basis of Chinese word segmentation, the dynamical searching and composition of Web services is implemented, so that the precision ratio and the recall ratio of Web services are increased.

Key words:Chinese word segmentation; Web service; OWL-S; UDDI; semantic retrieval

0 引 言

缺乏语义是现有 Web 服务的一个突出缺陷。Web 服务标准 UDDI 提供一种基于分布式商业注册中心机制来进行服务管理、服务注册和服务发现。这种方法的主要缺点是服务描述缺少灵活性、服务匹配采用简单的基于关键词搜索方法,这显然不能满足服务发现的智能化要求^[1]。对 Web 服务进行语义标注(semantic annotation)是解决问题的关键。

针对上述问题,文中通过对语义 Web 服务的

OWL-S 描述与相关技术的研究,将语义 Web 技术应用到 Web 服务过程中。由国外学术界提出的语义 Web 服务检索模型并没有充分考虑中西语言之间的差异,应用其进行中文检索时必须对查询请求进行中文分词处理,因此需要将语义 Web 服务 OWL-S 框架和中文分词技术相结合。

基于中文分词的语义 Web 服务 OWL-S/UDDI 检索模型,充分考虑了中文的语言特点,对中文检索请求语句进行中文分词和语义扩展,生成 Web 服务请求的 OWL-S 文档,继而实施语义检索匹配。

语义 Web 服务的研究框架主要有 WSDL-S、OWL-S、WSMO 三种。基于 OWL-S 的语义 Web 服务开发主要有两种途径^[2]:代码驱动(code-driven)和模型驱动(model-driven)。代码驱动的语义 Web 服务 OWL-S 框架主要以 Java 代码为起点,产生 OWL-S 描述本体,这种方式适应 Web 服务发现过程。

在检索过程中,传统的 OWL-S/UDDI 匹配模型

收稿日期:2008-05-04

基金项目:国家自然科学基金(60672031);辽宁省自然科学基金(20072142)

作者简介:牟 帅(1984-),男(满族),辽宁清原人,硕士研究生,CCF 会员,研究方向为智能信息处理;黄映辉,博士,教授,CCF 会员,研究方向为智能信息处理;李冠宇,硕士,教授,CCF 会员,研究方向为智能信息处理。

没有考虑中文词语间并无类似英文中空格的分隔标志,无法对中文检索的请求语句进行适当处理。根据语义 Web 服务 OWL-S 框架,基于 OWL-S 的语义 Web 服务匹配主要由四个部分构成:一是 Web 服务的 OWL-S 描述,并通过语义 Web 服务的 OWL-S 模型进行服务检索;二是语义 Web 服务检索语句的中文分词,提取关键词,并进行语义标注,结合所收集到的信息生成 OWL-S 请求文档;三是通过 OWL-S 与 UDDI 映射增强 UDDI;四是利用检索模型对 OWL-S 请求文档与 OWL-S/UDDI 注册中心的服务描述进行匹配从而进行检索。

1 语义 Web 服务的 OWL-S 描述

1.1 语义 Web 服务的 OWL-S 描述

马里兰大学信息及网络动态实验室(MIND)开发的 OWL-S API 可用来具体实现语义 Web 服务的描述转换,即将 WSDL 文档转换为 OWL-S 文档。OWL-S API 提供了一个 Java API 编程接口,用于读取、执行和编写服务的 OWL-S 描述,支持读取不同版本的 OWL-S,包括 DAML-S 0.7、OWLS 0.9、OWLS 1.0 等^[3]。在 URL 中输入相应的 WSDL 地址,解析可生成相应的 OWL-S 文档。

对于一个服务,若有一个操作,就生成一个 OWL-S 文档,文件名即为这个操作的名称;若有多个操作,就对应多个 OWL-S 文档。OWL-S API 还提供了一个执行引擎,既可以调用原子进程,还可以调用复合进程。原子进程带有 WSDL 或者 OWL-S 描述,复合进程可以使用控制结构,如顺序(sequence)、无序(split)和选择(choice)^[4]。另外,OWL-S API 里的所有对象都继承 OWL Resource 类,因而具有 accessor 函数,来获取 OWL 属性的值,也可获取各层的数据模型(基于 Jena)用于更高级的查询。

1.2 语义 Web 服务的 OWL-S 框架

应用代码驱动的语义 Web 服务的 OWL-S 框架时,开发者只需要编写代码来描述功能完善的 Web 服务,然后转化为语义描述。最重要的是,转化过程可通过相应工具自动完成,开发难度相对降低和实施过程比较容易,并减少了产生错误的可能。例如,通过 Axis 和 .Net Web 服务框架可自动地从 Web 服务的源代码生成 WSDL 描述文档;此外,服务的 OWL-S 描述也可以用 WSDL2OWL-S 工具自动生成。同样,Web 服务的 OWL-S 文档中 Service Profile 可以通过 OWL-S2UDDI 工具自动与 UDDI 的发布广告信息进行映射。语义 Web 服务的 OWL-S 框架如图 1 所示。

如图 1 所示,在语义 Web 服务 OWL-S 框架中,

Java 代码开发的 Web 服务,其描述用 WSDL 文档,在相应的领域本体支撑条件下转换成 OWL-S 本体,共同对 Web 服务进行描述,并将转换后的 OWL-S 本体信息添加进 OWL-S/UDDI 注册中心。当服务请求者发出请求时,需经过检索模型对查询请求进行处理,进而在注册中心进行匹配检索,匹配过程也可参考原有的 WSDL 服务描述,匹配结果经检索模型筛选后反馈给服务请求者进行服务调用。

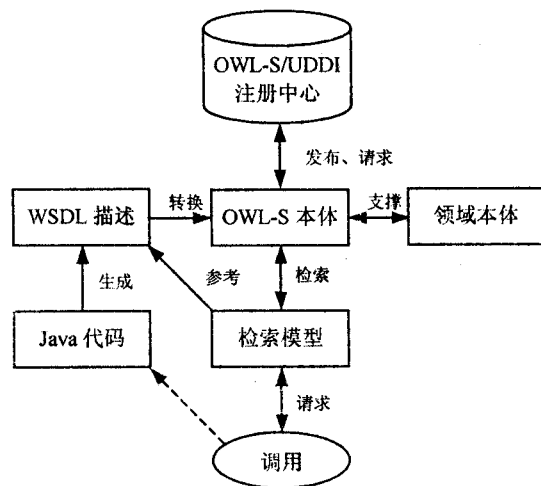


图 1 语义 Web 服务 OWL-S 框架

2 中文检索语句的处理

2.1 检索语句的中文分词

无论以什么方式描述 Web 服务,对其检索过程的输入始终同普通搜索一样使用的是自然语言,检索条件的输入并非只是关键词,还可以是合成词、词组甚至是句子,因此需对其进行中文分词,过滤得到所需的关键词。

无论是基于词典与规则的分词法还是基于统计的分词法,都离不开分词词典的辅助。用 WordNet、同义词林^[5]等来指导中文分词、语义标注、句法分析和语义分析的正确进行。使用中文分词算法结合统计规律和语义规则排除歧义,提高中文分词的精度。

Web 服务检索总是面向某个或某几个特定领域的。明确该领域范围,避免无关信息的噪声污染,为判别中文分词结果提供了依据,可提高查准率。

2.2 检索语句的语义标注

用户的注册信息直接通过动态网页存储于 Web 数据库中,而服务请求、访问记录、客户端(OS、IP 等)的信息由 IC-Agent(信息搜集代理)进行简单分类再存入 Web 数据库,具体检索过程是将检索请求和上述信息一同发送至 OWL-Agent(OWL-S 标注代理)进行语义标注,结合本体库、知识库等信息生成请求

OWL-S 文档。已有 OWL 标注器 Oiled、iAnnotate Tab 等工具可自动或半自动地完成语义标注工作^[6]。

在语义标注过程中,根据 OWL-S 的 Service Profile、Service Model、Service Grounding 三个本体,将信息分为三类:

(1) 请求服务功能概况。包括所需服务能提供的功能、返回结果类型等,由服务请求信息形成;

(2) 请求服务范围。包括服务所面向的用户类别、在具体领域中的范围,由用户注册信息、访问记录统计信息等形式;

(3) 请求服务适用环境。指 Web 服务间交互采用的通信方式、数据传送方式等,由客户端信息形成。

3 OWL-S 增强 UDDI

为利用语义匹配和 OWL-S 提供的基于功能的检索,需要将 OWL-S 的 Service Profile 信息嵌入到 UDDI 发布中。采用 OWL-S 本体来指导服务的注册、管理和预发现过程,可使对服务的描述更加准确。如果一个 Service Profile 元素有一个相应的 UDDI 元素,则采用一一对应的映射;如果没有相应的 UDDI 元素,就采用基于 tModel 的映射^[7]。

对于 Web 服务发现,可用 OWL-S 的 Service Profile 辅助 UDDI;对于 Web 服务组合,可将 OWL-S 的 Service Model 与现有的 BPEL4WS(business process execution language for Web services)和 WSFL(Web services flow language)等技术结合使用^[8]。Service Profile 的最大特点是双向性,服务提供者用它描述服务功能,服务请求者用它描述需求,服务发现过程利用这种双向信息进行匹配。此时 UDDI 可通过 Service Profile 提供功能检索,而不是传统的关键词检索。OWL-S 的功能描述采用 UDDI 中的 tModel 编码来实现。服务发现过程中还需要了解服务质量、服务分类等。Service Profile 通过 ServiceParameter 属性来具体描述 Web 服务的其它性能参数,通过 ServiceCategory 属性指向特定领域服务分类体系的详细条目。

4 语义 Web 服务的 OWL-S 检索模型

用 OWL-S 代码驱动方式来进行语义 Web 服务的开发。服务描述和服务发布广告都用 OWL-S 描述,语义 Web 服务的发现机制 OWL-S/UDDI 则以 OWL-S 为搜索条件。采用基于服务提供者、服务注册中心和服务请求者三者间交互的 SOA(Service-Oriented Architecture)体系结构和语义 Web 服务技术,涉及服务的发布、查找、组合和绑定操作^[9]。OWL-S 是连接语义 Web 和 Web 服务两大技术的桥梁。语义

Web 服务的 OWL-S 检索模型主要围绕 OWL-S 展开。具体模型如图 2 所示。

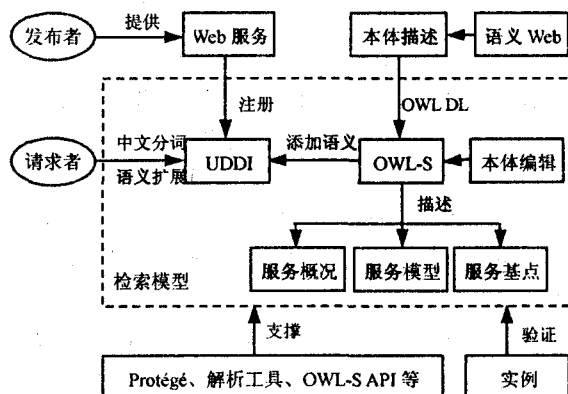


图2 语义 Web 服务的 OWL-S 检索模型

Web 服务注册于 UDDI。语义 Web 用 OWL-S 本体描述 Web 资源并为其添加语义信息。基于中文分词技术的检索模型,用 OWL-S 来描述 Web 服务的概况、模型和基点,对 UDDI 添加语义形成 OWL-S/UDDI 的服务发现和服务匹配机制,并对检索请求进行中文分词和语义标注处理,进而与 OWL-S/UDDI 中的服务的 OWL-S 描述进行语义匹配,达到语义检索的目的。服务的 WSDL 描述经 OWL-S API 转化成 OWL-S 描述,由于转换过程中会丢失数据流、工作流等信息,所以要对 OWL-S 本体进行编辑。

5 实验

5.1 实验过程

为验证上述模型的实用价值,通过实例和相应工具将普通的 UDDI 发布检索和经过 OWL-S 转换的 jUDDI^[10]发布检索进行对比。

使用的 UDDI 是基于 Java 的免费开源工具包 jUDDI,建立一个属于自己的 UDDI 注册中心,可以直接调用 jUDDI 发布程序。从 <http://www.mindswap.org>、<http://www.webservice.net> 和 <http://www.example.org/owls/> 等网站下载已有的 ZipService、AddressPhotoService、bnprice、books、TranslatorService 等 Web 服务注册到本机。若下载的服务是用 WSDL 描述的,通过 OWL-S API 将其转换成 OWL-S 描述。

若有服务请求者在 Web 上提出:“我要买计算机的书”,则语义 Web 服务请求的处理过程为:

(1) 服务请求的中文分词。采用第三代智能分词系统 3GWS^[11]进行中文分词并对所划分的各个单词标注词性,进而过滤剔除无用词汇。由于要通过 OWL 进行语义标注,所以省去简单语义扩展过程。之所以采用 3GWS 进行中文分词,是因为 3GWS 是免费下载且提供开源的输入(TGWS_Init)、输出(TGWS_Ex-

it)、用户输入路径(TGWS_ImportUserDict)等 API 函数接口的中文分词系统。本实验的服务请求的中文分词结果为“我/rr 要/v 买/v 计算机/n 的/rdel 书/n”。

(2) 服务请求的语义标注。为了理解服务请求者检索请求的语义,需对中文分词结果进行语义标注,用 OWL-S 来描述服务请求,带有语义信息。采用 Protégé 和 iAnnotate Tab^[12]来进行标注,iAnnotate Tab 是一个基于 OWL 的 Protégé 插件,手动用本体概念对纯文本或 Word 文档进行标注,不同的概念用不同的颜色表示。

(3) 语义 Web 服务匹配调用。采用 OWLS-MX 检索一系列用 OWL-S 描述的服务,其提供的图形用户界面可以引导完成匹配操作的全过程^[13];语义 Web 服务的调用需要输入的信息是:源文件路径、OWL-S 描述文档、jUDDI 注册中心等,根据匹配引擎返回的服务的 OWL-S 描述以及服务请求所携带的相关 IOPE 参数信息等被传递给服务调用模块(WSIF),进而向目标服务发出调用请求,并将最终结果返回到服务请求模块,反馈给服务请求者。

5.2 实验结果

分别进行关键词匹配和应用语义 Web 服务 OWL-S 检索模型通过 OWL-S VM 的语义匹配,对比二者的检索效果。检索效果评价指标为查准率和查全率。查准率=检索到的符合检索请求的文档数/检索的总文档数,查全率=检索到的与检索请求相关的文档数/检索的总文档数。计算得出具有相对意义的实验数据如表 1 所示。

表 1 关键词检索和语义检索的对比

注册方式	关键词检索		语义检索	
	查准率	查全率	查准率	查全率
UDDI	0.30	0.20	-	-
jUDDI	0.30	0.20	0.75	0.60

实验结果表明,UDDI 注册的 Web 服务只支持关键词匹配,而 jUDDI 注册的语义 Web 服务既可以用关键词检索也可以用语义检索。两种注册方式的关键词检索的结果相同,查准率 0.33,查全率 0.20。采用 jUDDI 和语义扩展,对 Web 服务检索的性能有明显提升,查准率为 0.75,查全率也上升了 0.40。查全率的数值表明,通过语义扩展可以检索到关键词不匹配但确为用户所需的服务,甚至是组合服务,相比关键词匹配方式有显著的提高。

6 结束语

通过对 Web 服务的 OWL-S 描述及 OWL-S 框

架的分析,结合中文分词、OWL-S 技术,构建了基于中文分词的语义 Web 服务 OWL-S 检索模型。该模型从检索请求和注册中心两个方面着手:对检索请求采用中文分词技术提取关键词,结合收集到的信息对其进行语义标注,生成检索请求的 OWL-S 文档;对于 UDDI,通过与 OWL-S 的映射可实现对检索请求的 OWL-S 文档的匹配,进而由虚拟机匹配得出结果。

实验数据表明,通过中文分词后,该模型所述方法有利于提高检索的查准率和查全率,取得了语义 Web 服务检索的效果。

参考文献:

- [1] Sheila A, Martin D L. Bring Semantics to Web Services[J]. IEEE Intelligent Systems, 2003,18(1):90-93.
- [2] Srinivasan N, Paolucci M, Sycara K. CODE: A Development Environment for OWL-S Web Services [EB/OL]. [2005-10-04]. <http://iswc2004.semanticweb.org/demos/10/paper.pdf>.
- [3] 高继峰. 基于语义 Web 的智能信息检索系统研究[D]. 郑州: 郑州大学, 2006.
- [4] 秦雪杰. 基于语义 Web 服务的业务过程集成研究与应用[D]. 南京: 河海大学, 2006.
- [5] 程涛,施水才,王霞,等. 基于同义词林的中文文本主题词提取[J]. 广西师范大学学报:自然科学版, 2007, 25(2):145-148.
- [6] 俞扬信. 基于 OWL-S 服务匹配的信息查询模型[J]. 计算机与应用化学, 2007,24(9):1277-1280.
- [7] 郭李娟,史浩山,王巍. 基于 OWL-S/UDDI 的 WEB 服务发现机制[J]. 科学技术与工程, 2007,7(9):2103-2107.
- [8] 梁晟. 基于语义 Web 的服务自动组合技术的研究[D]. 北京: 中国科学院, 2004.
- [9] 高岩,张少鑫,张斌,等. 基于 SOA 架构的 Web 服务组合系统[J]. 小型微型计算机, 2007,28(4):729-733.
- [10] Apache Web Services Project. Apache jUDDI [EB/OL]. [2007-12-11]. <http://ws.apache.org/juddi/>.
- [11] Fajava (Beijing) Intelligence Ltd. 3GWS: 3rd Generation Word Segmenter [EB/OL]. [2005-01-01]. <http://www.fajava.com>.
- [12] Chintan P. iAnnotate Tab [EB/OL]. [2005-10-08]. <http://www.dbmi.columbia.edu/cop7001/iAnnotateTab/iannotate.htm>.
- [13] Benedikt F, Klusch M. Hybrid OWL-S Web Service Matchmaker [EB/OL]. [2006-09-01]. <http://www-ags.dfki.uni-sb.de/klusch/owls-mx/>.