

基于相似度的语义Web服务发现技术研究

陈文字, 张忠全, 向涛, 桑楠

(电子科技大学计算机科学与工程学院 成都 610054)

【摘要】针对传统Web服务在服务发现中存在效率低的问题,应用语义Web技术和本体理论,提出了一种基于相似度的语义Web服务发现模型。采用Web本体语言(OWL-S)描述Web服务,发布服务时包含充分的语义信息,在服务搜索中加入本体推理过程,综合度量服务的功能匹配程度和信誉值。实验结果表明,语义Web服务发现方法的查准率和查全率高于UDDI基于关键字的查找方法。

关键词 本体; 信誉; 语义Web; 服务发现; Web服务

中图分类号 TP311

文献标识码 A

doi:10.3969/j.issn.1001-0548.2010.06.019

Research on Similarity-Based Semantic Web Services Discovery

CHEN Wen-yu, ZHANG Zhong-quan, XIANG Tao, and SANG Nan

(School of Computer Science and Engineering, University of Electronic Science and Technology of China Chengdu 610054)

Abstract Aiming at the problems of low efficiency in service discovery of traditional web service, a similarity-based semantic web service discovery model is proposed, with the application of semantic web technology and ontology. The mechanism uses Ontology Web Language for Service (OWL_S) to describe web service and publishes service with semantic information. Ontology reasoning is added into service searching to measure its degree of functional matching and reputation value. The experiments suggest that the proposed method is better than that of Universal Description Discovery and Integration (UDDI) in precision and recall ratio.

Key words ontology; reputation; semantics Web; service discovery; Web services

Web服务是一种面向服务架构的技术,通过标准的SOAP协议提供服务,保证分布式和异构环境下的应用可以互操作。Web服务具有跨平台、复用性高、成本低和使用简单等优点,所以其应用范围非常广泛,特别是在电子商务领域^[1]。但其也面临一些挑战,如怎样从庞大的Web服务库中准确地搜索到所需的服务,即服务发现等。传统的Web服务将描述服务信息的WSDL发布到UDDI,通过关键字匹配的方式查找服务^[2-3]。该服务的发布和发现方式不支持基于语义的模糊匹配,不能选出与检索词同义和相关的服务,不利于实现服务的自动组合,难以构建高效、动态的系统。语义Web服务技术的诞生解决了传统服务发现的局限性,具有更好的信息表达能力和语义推理能力。目前,对语义Web服务的研究取得了部分进展^[4],但对语义Web服务发现机制的研究尚处于起步阶段。

本文引入语义Web服务技术,比较了传统Web服务和语义Web服务;引入本体理论,采用OWL_S替代WSDL描述带语义的Web服务;分析了文献[5]中的服务发现算法的不足,提出了基于功能相似度和信誉度的服务发现方法,给出了影响功能和信誉的各个因素以及它们之间的联系,并通过实验说明了该方法的可行性。

1 语义Web服务概述

1.1 语义Web服务的提出

互联网应用的快速发展使现有技术的局限逐渐显现出来。当前互联网技术没有描述信息的含义,也不关心信息的具体内容,导致信息处理的智能化水平较低^[6]。为改变该状况,文献[7-8]提出了语义Web技术。语义Web通过在Web资源中加入语义信息,扩展当前Web的自描述能力。在Web服务中引入语义

Web技术, 可以解决传统服务中不含语义信息, 计算机之间不能相互理解和操作的问题, 并能将Web服务的功能和信誉信息转化为格式化的语义信息, 方便机器的自动识别和匹配, 提高服务查找的精确度。

语义Web服务是语义Web和Web服务两种技术的融合, 是Web服务发展的新方向。表1给出了语义Web服务与传统Web服务相比较的优点^[9]。

表1 传统Web服务和语义Web服务的比较

| 项目 | 服务 | 服务请求者 | 服务描述 | 数据交换 |
|---------|-----|-------|------|------|
| 传统Web服务 | 简单的 | 人 | 基于词汇 | 基于语法 |
| 语义Web服务 | 复合的 | 人或代理 | 基于本体 | 基于语义 |

1.2 基于本体的语义Web服务

语义Web建立在本体论的基础上, 本体定义了事物的属性、约束、类以及类之间的关系等信息, 在Web服务中引入本体可以增加描述服务的语义信息。由于本体具有清晰的概念层次结构, 并支持逻辑推理, 所以可以检测请求服务和待选服务的一致性, 使检索出来的信息更符合要求的, 增强服务自动发现的能力并提高服务查全率。为了进一步达到服务的自动发现和执行, 还需要加强对Web服务语义信息的描述, 因此催生了OWL_S。OWL_S是基于OWL的Web服务本体, 其标记语言结构准确地刻画了Web服务的功能和属性。OWL_S不仅支持基于语义的推理机制, 而且对信息的表达能力也比WSDL强, 其主体结构如图1所示^[10]。

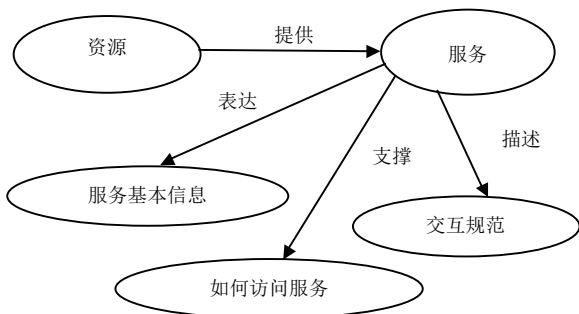


图1 OWL_S主体结构

从图1可以看出, Service Model、Service Gounding和Service Profile是OWL_S的主体结构的3个组成部分, 它们分别通过属性describedBy、supports和presents描述Web服务的语义信息, 包括质量语义、数据语义和功能语义等, 这些语义信息组合起来为服务发现提供依据。

2 基于相似度的语义Web服务发现

文献[5]提出了一种基于功能相似程度进行匹配

的服务发现方法, 其中评价服务请求和待选服务之间的功能相似程度是通过它们之间的公有信息进行的。分别计算OWL_S对象的Service Gounding、Service Profile和Service Model以决定待选服务是否匹配服务请求, 但该发现方法存在以下缺点^[11-12]:

- (1) 需要参照Service Gounding、Service Profile和Service Model的信息, 计算比较复杂, 模型实现困难;
- (2) 筛选服务时没有考虑其信誉度的因素, 可能导致某些信誉低下, 没有达到指定阈值的服务也被筛选出来;
- (3) 性能和查全率低, 尤其在海量服务库中查找时比较明显。

为解决以上问题, 本文采用本体推理理论和语义Web技术, 提出一种新的服务发现方法。本文的服务发现主要关注服务的功能语义相似程度和服务提供方的信誉值, 并加入性能度量, 不再考虑时效性和安全等其他非功能性因素。

假定服务请求和服务公告分别用request和advertise表示, 它们的相似度用similarity(request, advertise)表示, 请求服务和待选服务之间的功能相似度和信誉度相似度分别用similarity_{function}(request, advertise)和similarity_{reputation}(request, advertise)表示。请求服务和待选服务的相似度定义为:

$$\begin{aligned} \text{similarity}(\text{request}, \text{advertise}) = & C_1 \times \text{similarity}_{\text{function}}(\text{request}, \text{advertise}) + \\ & C_2 \times \text{similarity}_{\text{reputation}}(\text{request}, \text{advertise}) \end{aligned} \quad (1)$$

并且 $\sum_{i=1}^2 C_i = 1$ 。

2.1 功能相似度

输入、输出、前置条件和效果是影响功能相似度的4个主要因素(简称IOPE)。服务请求和服务公告的功能相似度定义为:

$$\begin{aligned} \text{similarity}_{\text{function}}(\text{request}, \text{advertise}) = & C_1 \times \text{similarity}_{\text{input}}(\text{request}, \text{advertise}) + \\ & C_2 \times \text{similarity}_{\text{output}}(\text{request}, \text{advertise}) + \\ & C_3 \times \text{similarity}_{\text{precondition}}(\text{request}, \text{advertise}) + \\ & C_4 \times \text{similarity}_{\text{effect}}(\text{request}, \text{advertise}) \end{aligned} \quad (2)$$

并且 $\sum_{i=1}^4 C_i = 1$ 。

2.2 信誉度相似度

可用性、响应时间和价格是影响服务信誉度的3个主要因素。分别用similarity_A、similarity_T和similarity_P表示可用性相似度、响应时间相似度和价格相似度。信誉度相似度为:

$$\text{similarity}_{\text{rep}} = C_1 \times \text{similarity}_A + C_2 \times \text{similarity}_T + C_3 \times \text{similarity}_P \quad (3)$$

并且 $\sum_{i=1}^3 C_i = 1$ 。

2.3 服务发现算法

服务发现的具体步骤如下:

(1) 遍历服务库中登记的所有Web服务的公告,通过语义推理过滤与请求服务在功能上不相似的服务,得到有效的待选服务集合;

(2) 依次查找有效的待选服务,筛选淘汰信誉值低下的服务;

(3) 计算各个待选服务的功能语义相似度;

(4) 计算各个待选服务的信誉度相似度;

(5) 计算整体相似度,并按照相似度从大到小的顺序排列;

(6) 返回相似度达到指定阈值的列表,并选择相似度最大的服务作为最佳匹配的服务。

3 应用实例

3.1 定义本体

以查找汽车销售服务为例说明服务匹配过程。假定服务注册库中有SA销售自行车、SB售火车票、SC销售轿车、SD只销售本田汽车、SE销售所有类型的车等5个满足本体规范的有效服务。定义如下的本体实例:

```

<daml:DataProperty rdf:about= "#汽车">
  <daml:SamePropertyAs>
    <daml:class rdf:about= "automobile" />
  <daml:SamePropertyAs/>
  <daml:subPropertyOf>
    <daml:Class rdf:about= "#交通工具" />
  <daml:subPropertyOf />
  <rdfs:domain>
    <rdf:name= "SA" value= "bicycle" />
    <rdf:name= "SB" value= "tickets" />
    <rdf:name= "SC" value= "car" />
    <rdf:name= "SD" value= "honda" />
    <rdf:name= "SE" value= "vehicle" />
  <rdfs:domain />
</daml:DataProperty />

```

3.2 Web服务匹配

注册服务中的SB销售火车票,与请求服务-汽车销售在概念和功能上不一致,所以经过语义推理分析,排除该项服务,剩余的4个服务都是有效的待选服务。

3.2.1 计算功能相似值

影响功能相似度的输入、输出、前置条件和效果等参数如表2所示,假设输入对于所有的服务都是相同的。

表2 功能参数等级表

| 服务 | 输入 | 输出 | 前置条件 | 效果 |
|----|----|-----|------|-----|
| SA | 1 | 0.6 | 0.7 | 0.5 |
| SC | 1 | 0.9 | 0.9 | 0.8 |
| SD | 1 | 0.7 | 0.6 | 0.7 |
| SE | 1 | 0.7 | 0.8 | 0.6 |

假定输入、输出、前置条件和效果的权重分别为0.3、0.3、0.2、0.2。根据式(2),采用加权平均值计算出各个服务的功能相似值为:

$$SA=0.3 \times 1 + 0.3 \times 0.6 + 0.2 \times 0.7 + 0.2 \times 0.5 = 0.72$$

$$SC=0.3 \times 1 + 0.3 \times 0.9 + 0.2 \times 0.9 + 0.2 \times 0.8 = 0.91$$

$$SD=0.3 \times 1 + 0.3 \times 0.7 + 0.2 \times 0.6 + 0.2 \times 0.7 = 0.77$$

$$SE=0.3 \times 1 + 0.3 \times 0.8 + 0.2 \times 0.8 + 0.2 \times 0.6 = 0.82$$

3.2.2 计算信誉度相似值

服务SA、SC、SD和SE的响应时间、可用性、价格等信誉参数如表3所示,并且它们的权重分别为0.5、0.3和0.2。根据式(3),计算出各个服务的信誉值为:

$$SA=0.5 \times 0.6 + 0.3 \times 0.8 + 0.2 \times 0.8 = 0.7$$

$$SC=0.5 \times 0.8 + 0.3 \times 0.9 + 0.2 \times 0.7 = 0.81$$

$$SD=0.5 \times 0.5 + 0.3 \times 0.8 + 0.2 \times 0.6 = 0.61$$

$$SE=0.5 \times 0.6 + 0.3 \times 0.9 + 0.2 \times 0.8 = 0.73$$

假设信誉的阈值为0.65,即认为低于该信誉值的服务包含有虚假信息,不将其添加到待选服务列表中,故丢弃服务SD。

表3 信誉等级表

| 服务 | 响应时间 | 可用性 | 价格 |
|----|------|-----|-----|
| SA | 0.6 | 0.8 | 0.8 |
| SC | 0.8 | 0.9 | 0.7 |
| SD | 0.5 | 0.8 | 0.6 |
| SE | 0.6 | 0.9 | 0.8 |

3.2.3 整体语义相似度

假定功能相似度和信誉相似度的权重分别是0.7和0.3,根据式(1),计算出各个服务的整体相似度分别为:

$$SA=0.7 \times 0.72 + 0.3 \times 0.7 = 0.714$$

$$SC=0.7 \times 0.91 + 0.3 \times 0.81 = 0.92$$

$$SE=0.7 \times 0.82 + 0.3 \times 0.73 = 0.793$$

综合考虑功能相似度和信誉度, 得出各个服务的整体相似度。各服务相似度从大到小的排序以次是SC、SE、SA。所以对本例而言, 应该首选SC作为最佳匹配的服务。

4 试验验证与性能分析

4.1 实验验证

仍然用汽车销售实例验证前文所提模型的可行性。实验环境配置为本体建模工具Protege3.1、本体编辑工具OWL_S Editor、推理机Jena+RACER Pro1.9、私有注册中心jUDDI 0.9rc3、应用服务器Jakarta Tomcat5.0、数据库服务器MySQL5.1和开发平台Eclipse。

先编写相应的服务并在UDDI中注册, 如图2所示。

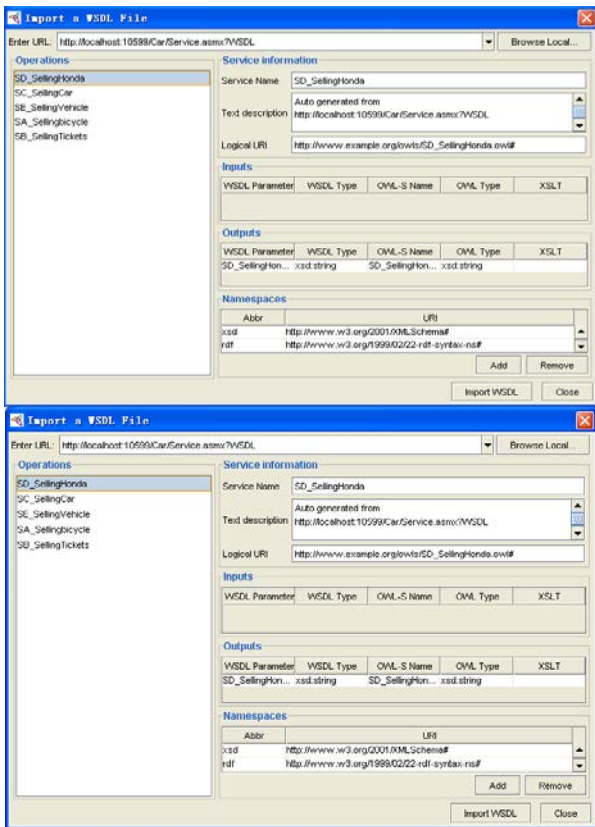


图2 Web服务注册



图3 Web服务的查询结果

图2左侧窗口中列出了5个与汽车销售相关服务, 右侧窗口是对这些服务信息的详细描述。要得到所需Web服务列表, 用户需要提供必要的信息如Service Name、Description、Inputs和Outputs等给请求代理, 代理根据语义推理和信誉度量过滤淘汰不相关和虚假的服务, 然后按服务相似度匹配程度返回满足要求的Web服务列表。如图3所示。

4.2 性能分析

目前还没有统一的计算Web服务发现性能的标准, 本文将采用查准率和查全率作为评价依据。

由于本文实验的服务集数量有限, 不能有效地验证服务发现性能。为收集大量的试验数据, 从uddi.ibm.com等站点下载了110个描述Web服务的WSDL文件和40个OWL_S文件。服务主要有汽车销售、股票行情、航班查询和数字图书馆查询服务等。用工具OWL_S Editor对WSDL文件进行了转化, 得到150个OWL_S文件, 然后分别基于UDDI、Semantic进行100次有效的服务请求, 服务匹配结果如图4所示。

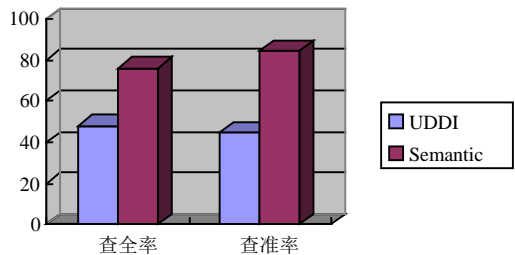


图4 两种方法查全率和查准率的比较

图4表明, 基于关键字查找的查准率和查全率(48%,41%)远低于基于OWL_S语义相似度查找的查准率和查全率(78%,85%), 原因是UDDI基于关键字的服务发现机制不支持自动化的服务发现, 查询结果需要人工处理, 才能确定服务是否和查询条件相匹配。而基于语义的Web服务发现机制能够自动查找并返回与所要查找内容在功能和信誉方面达到指定相似度的Web服务。

5 结束语

传统的基于关键字和简单分类的Web服务发现存在查准率和查全率底下的缺点。本文提出了一种基于相似度的语义Web服务发现方法, 该方法在服务发现过程中增加了语义信息, 通过语义推理增强服务的信誉度相似度和功能相似度, 从而提高服务发现和匹配的精确程度。实验验证了该方法的可行性和相对于传统方法的优越性。进一步的研究包括充分利用本体推理机制优化语义Web服务发现算法, 并提高语义Web服务发现的性能。

(下转第910页)

体制,同时为了获得更好的安全性选用了安全性较高的标准模型,所以本文的基础更加自然,安全性更可靠。如何在保证安全性的前提下提高效率是下一步的工作目标。

参 考 文 献

- [1] AL-RIYAMI S S, PATERSON K. Certificateless public key cryptography[C]//Advances in Cryptology Asiacrypt 2003. Taipei: Springe Verlag, 2003: 452-473.
- [2] GENTRY C. Practical identity-based encryption without random oracle[C]//Advances in Cryptology Eurocrypt 2006. Saint Petersburg: Springe Verlag, 2006: 445-464.
- [3] AU M H, CHEN J, LIU J K, et al. Malicious KGC attack in certificateless cryptography[C]//ACM Symposium on Information, Computer and Communications Security 2006. Taipei: ACM Press, 2007.
- [4] HUANG Q, WONG D S. Generic certificateless encryption in the standard model[C]//Advances in Information and Computer Security, IWSEC 2007. Nara: Springe Verlag, 2007: 278-291.
- [5] DENT A W, LIBERT B, PATERSON K G. Certificateless encryption schemes strongly secure in the standard model[C]//11th International Conference on Public Key Cryptography. Barcelona: Springe Verlag, 2007.
- [6] ONG H, CHOI K Y, HWANG J Y, et al. Certificateless public key encryption in the selective-ID security model (Without Random Oracles) [C]//Pairing-Based Cryptography-Pairing 2007. Tokyo: Springe Verlag, 2007: 60-82.
- [7] BAEK J, SAFAVI, NAINI R, SUSILO W. Certificateless public key encryption without pairing[C]//Information Security. Singapore: Springe Verlag, 2005: 134-148.
- [8] CHENG Zhao-hui, CHEN Li-qun, LING Li, et al. General and efficient certificateless public key encryption constructions[C]//Pairing-Based Cryptography-Pairing 2007. Tokyo: Springe Verlag, 2007: 83-107.
- [9] WATERS B. Efficient identity-based encryption without random oracles[C]//Advances in Cryptology EUROCRYPT 2005. Aarhus: Springe Verlag, 2005: 114-127.
- [10] DENT A W. A survey of certificateless encryption schemes and security models[J]. International Journal of Information Security, 2008,7(5): 349-377.
- [11] CHENG Z, COMLEY R. Efficient certificateless public key encryption[R/OL]. [2009-03-14]. <http://eprint.iacr.org/2005/012/>

编辑 黄 莘

(上接第899页)

参 考 文 献

- [1] FEI Chun, TANG Xue-fei. Research on the E-learning application of web service[J]. Journal of Electronic Science and Technology of China, 2005, 3(3): 218-221.
- [2] 郑向宏, 李院春, 李增智, 等. 面向语用Web服务的Qos评价模型研究[J]. 电子科技大学学报, 2007, 36(6): 1477-1480.
ZHENG Xiang-hong, LI Yuan-chun, LI Zeng-zhi, et al. Research on pragmatic web-oriented Qos evaluation model [J]. Journal of University of Electronic Science and Technology of China, 2007, 36 (6): 1477-1480.
- [3] 赵文峰, 孟祥武, 陈俊亮. 信息提供类Web服务与RDF数据源的集成[J]. 北京邮电大学学报, 2008, 31 (6): 109-112.
ZHAO Wen-feng, MENG Xiang-wu, CHEN Jun-liang. Integration of information-providing web services and RDF data sources[J]. Journal of Beijing University of Posts and Telecommunications, 2008, 31(6). 109-112.
- [4] PAPAOGLOU M P, GEORGAK O D. Service oriented computing[J]. Communication of the ACM, 2003, 46 (10): 25-28.
- [5] DAY J, DETERS R. Selecting the best web service[C]//Proceedings of the 2004 Conference of the Centre for Advanced Studies on Collaborative Research Tabel of Contents. Markham: [s.n.], 2004.
- [6] 林清滢. 基于UDDI的语义Web服务发现研究[J]. 计算机工程与设计, 2006, 27(12): 2215-2217.
LIN Qing-ying. UDDI based semantic web service discovery research[J]. Computer Engineering and Design, 2006, 27(12): 2215-2217.
- [7] KAMVAR S D, SCHLOSSER M T. In: Reputation management in P2P networks[C]//Proc of the 12th World Wide Web Conference. Hawaii: ACM Press, 2004: 123-134.
- [8] RAMA A, RICHARD C. A method for semantically enhancing the service discovery capabilities of transaction on web service, 2003, 3(3): 310-323.
- [9] 艾未华, 宋自林, 魏 磊, 等. 基于领域本体的Web服务发现[J]. 电子科技大学学报, 2007, 36(3): 506-509.
AI Wei-hua, SONG Zi-lin, WEI Lei, et al. Web service discovery based on domain ontology[J]. Journal of University of Electronic Science and Technology of China, 2007, 36(3): 506-509.
- [10] 王 慧, 王金华, 赵煜辉, 等. 基于信誉的语义Web服务发现[J]. 计算机科学, 2007, 34(8): 130-134.
WANG Hui, WANG Jin-Hua, ZHAO Yu-Hui, et al. Reputation-based semantic web service discovery[J]. Computer Science, 2007, 34(8): 130-134.
- [11] CURBERA F, KHALAF R, MUKHI N. The next step in web service[J]. ACM, 2003, 46(10): 29-34.
- [12] 白东伟, 刘传昌, 陈俊亮. 一种增强语义精确度的Web服务匹配方法[J]. 北京邮电大学学报, 2006, 29(5): 40-44.
BAI Dong-wei, LIU Chuan-chang, CHEN Jun-liang. A web services matchmaking method with enhanced semantic precision[J]. Journal of Beijing University of Posts and Telecommunications, 2006, 29(5): 40-44.

编辑 漆 蓉