

基于 OWL-S 的语义 Web 服务发现方法

刘杰, 傅秀芬

(广东工业大学 计算机学院, 广东 广州 510006)

摘要:随着 Web 服务的广泛应用和迅速发展,服务的种类和数量越来越多,要在众多的 Web 服务中高效地发现满意的服务成为一个关键的问题。目前的 Web 服务发现方法大多是采用基于关键字匹配的机制,查全率和查准率不高,难以满足用户的需要。针对这个问题将 OWL-S 引入服务发现,为 Web 服务添加语义信息,运用本体推理技术,文中提出一种基于 OWL-S 的语义相似度匹配度计算的 Web 服务发现方法,并设计了服务发现原型系统。实验结果表明提出的服务发现方法能够获得较好的服务发现效果。

关键词:Web 服务;服务发现;OWL-S;语义相似度

中图分类号:TP31

文献标识码:A

文章编号:1673-629X(2012)04-0073-04

A Method of Semantic Web Service Discovery Based on OWL-S

LIU Jie, FU Xiu-fen

(Faculty of Computer, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: With the rapid development and extensive use of Web services, the type and quantity of services is increasing, to find a satisfactory and efficient service among a large number of Web services has become a key issue. The current Web service registration and matching process has led to low recall and precision for service discovery and has much potential improvement for their performance. Propose a Web service discovery method utilizing semantic matching degree calculation. It described the Web service by using OWL-S to strengthen semantic information, and add semantic information to Web services, using ontology reasoning technology, and then the similarity degree of Web service is calculated, at last the request service matched the target service by using the similarity degree. Relative algorithms and the Web service discovery prototype system are designed. The experimental results demonstrate that proposed method can achieve better service discovery effectiveness.

Key words: Web service; service discovery; OWL-S; semantic similarity degree

0 引言

Web 服务作为一种新型的分布式计算技术,它是一种基于网络环境下的自适应、自包含、自描述、模块化的应用程序,促进了软件体系结构从基于实现技术的紧耦合、严格、静态的解决方案向基于商务模型的松耦合、灵活、动态的解决方案发展^[1,2]。

随着 Web 服务的迅速发展和广泛应用,Web 服务的数量也逐渐增多,要在大量的 Web 服务中有效地发现用户满意的服务是目前迫切需要解决的重要课题^[3]。目前存在的 Web 服务大多采用 Web 服务描述语言 WSDL 来描述服务, WSDL 只是在语法层面上对服务进行简单地描述,没有对服务的功能性等关键信

息进行语义标注,计算机无法处理和理解,服务发现方法是采用基于关键字匹配的机制,会漏掉许多和查询概念意义相同或相关的服务,服务发现的效率不高^[4,5]。为了解决上述问题,学术界在 Web 服务中应用语义 Web 技术,提出基于语义的 Web 服务发现方法。在 Web 服务中引入领域本体可以增强描述服务的语义信息,采用语义 Web 的知识标记方法描述 Web 服务,从而使计算机能够理解和分析 Web 服务信息,使得服务的自动发现、匹配、执行和组合成为可能^[6-8]。与传统的基于 WSDL 的服务发现方法相比,基于语义的 Web 服务发现方法更能满足用户的需要,服务发现的效率有明显的改善。文中采用 OWL-S 描述 Web 服务和服务请求信息,提出了一个基于本体概念相似度匹配的服务发现方法。

1 语义 Web 服务概述

1.1 语义 Web 服务

语义 Web 服务是 Web 服务技术和语义 Web 技术

收稿日期:2011-09-01;修回日期:2011-12-05

基金项目:广东省自然科学基金(9151009001000007)

作者简介:刘杰(1986-),男,硕士研究生,主要研究方向为协同软件和数据库技术;傅秀芬,教授,硕士生导师,主要研究方向为协同软件、数据库技术等。

的结合,语义 Web 是 Tim Berners-Lee 在 1998 年提出的一个概念,它是通过给万维网上的文档添加能够被计算机所理解的语义,从而使互联网成为一个通用的信息交换媒介,所以可以采用语义 Web 技术来描述 Web 服务,用户和软件代理可以理解服务的语义信息并能进行相应的推理,这样就使得服务的发现、选择、组合和执行的自动化成为可能^[9,10]。

1.2 Web 服务本体描述语言 OWL-S

OWL-S (Ontology Language for Web Services) 是基于 Web 本体语言 (OWL) 的 Web 服务本体,是把 Web 服务和语义 Web 技术相结合,它由 DAML-S 发展而来。OWL-S 定义了一组核心的语言标记体系,可以对 Web 服务进行语义上的描述,由它描述的 Web 服务能够被计算机准确地分析和理解,这样就使得 Web 服务的自动发现、调用、组合和监控成为可能。OWL-S 对 Web 服务的描述主要包括以下三个方面:

服务轮廓 (ServiceProfile): 描述服务是干什么的,也就是对服务的整体情况进行一个概括描述,比如服务提供者的信息、服务的功能可以解决什么样的问题,还有服务其它的一些特征信息等;

服务模型 (ServiceModel): 描述服务是如何工作的,包括服务执行的先后顺序、服务执行的过程以及调用服务所依照的流程等,据此服务请求者可以按照流程约束将多个服务组合起来构成更复杂的服务,并且能监控服务的运行情况;

服务绑定 (ServiceGrounding): 描述如何使用服务,包括具体的绑定信息和调用服务的具体细节信息,比如通讯采用协议的类型、所用消息的格式以及服务的寻址方式等。

OWL-S 的顶层结构如图 1 所示:

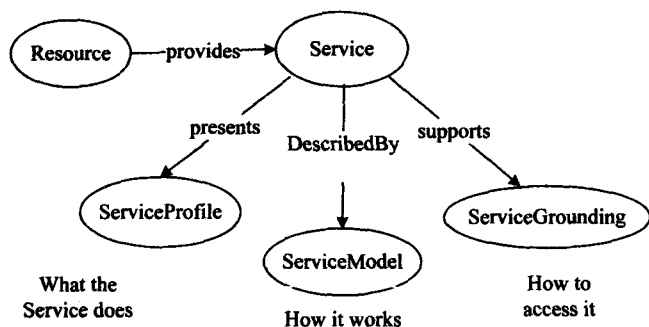


图 1 OWL-S 的顶层结构图

2 基于 OWL-S 的 Web 服务发现方法

2.1 语义 Web 服务相似度计算

服务匹配是基于语义的 Web 服务发现的关键,也就是如何计算服务发布信息和服务请求信息的匹配程度。所谓匹配度可以看做是用户请求和服务之间的相

似度,相似度越大说明服务越满足用户的需求,语义 Web 服务和请求的输入参数和输出参数都是用领域本体概念表示的,可以分别计算服务的输入本体和服务请求的输入本体语义相似度,与服务的输出本体和服务请求的输出本体语义相似度,对这两个相似度加权求出总的相似度,其关键为计算领域本体概念之间的语义相似度^[11,12]。在同一个领域本体框架中的两个本体概念的关系一般存在下面 3 种情况,如图 2 所示:

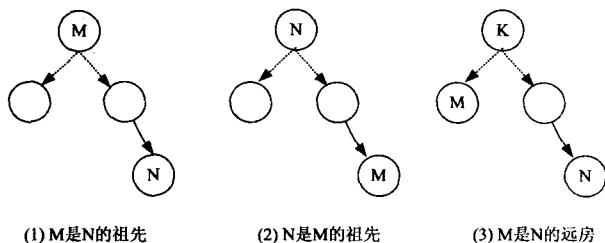


图 2 本体概念关系

在图 2 中,有这样的假设:首先,领域本体 M 和 N 之间的相似度是不对称的,即图 2(1)中 M 和 N 的相似度与图 2(2)中 N 和 M 的相似度不等价,且认为(1)中 M 与 N 的相似度要大于(2)中 M 与 N 的相似度;其次,远房的关系比不上嫡系的关系,比如图 2 中的(1)和(3),在(3)中 M 与 N 的关系是远房关系,而在(2)中 M 和 N 的关系是嫡系关系,(3)中 M 与 N 的相似度要小于(2)中 M 与 N 的相似度。此外,还有下面的情形,如图 3 所示:

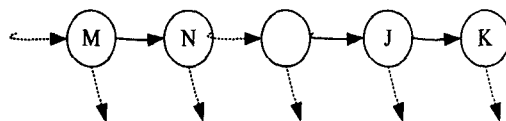


图 3 本体概念关系

很明显,在图 3 中,随着领域本体概念的细化,在本体下层的各个概念之间的差异会越来越小,M 和 N 之间的相似度大于 J 和 K 之间的相似度。

根据上面的分析,文中得出影响两个本体概念之间相似度的因素如下:

- (1) 本体在领域本体概念层次图中所在的深度大小;
- (2) 在同一领域本体概念层次图中的两个本体的关系;
- (3) 在领域本体概念层次图中的一条路径上的两个本体的距离。

设计了以下领域本体概念相似度度量函数:

$$\text{sim}(M, N) = \begin{cases} \frac{\text{deep}(M)}{\text{dis}(N, M) * \text{deep}(\text{Ontology})} \\ \frac{\text{deep}(N)}{\text{dis}(N, M) * \text{deep}(\text{Ontology})} \\ \frac{\text{deep}(L)}{(\text{dis}(M, L)^2 + \text{dis}(N, L)^2) * \text{deep}(\text{Ontology})} \end{cases} \quad (1)$$

上面的公式是计算本体 M 和 N 之间的相似度, $\text{deep}(\text{ontology})$ 为本体概念层次结构图中最大路径长度, $\text{dis}(M, N)$ 表示本体 M 和 N 之间的路径长度。服务请求和服务都采用 OWL-S 描述, 服务请求或服务中的输入参数可能有多个, 即用多个本体概念来表示, 同理输出也是一样的, 可以把它们看作是本体概念的集合。假设有两个本体概念集合 $O_x = \{O_1, O_2, \dots, O_m\}$ 和 $O_y = \{O_1, O_2, \dots, O_n\}$, 定义它们的本体概念相似度度量函数如下:

$$\text{sim}(O_x, O_y) = \prod_{i=1}^n \max\{s \mid s = \text{sim}(O_x, O_y) O_i \mid O_i \in O_x, O_j \in O_y, 1 \leq i \leq m, 1 \leq j \leq n\} \quad (2)$$

2.2 输入输出匹配

服务匹配是分别对服务请求的输入和输出的本体概念集合与候选服务对应的输入和输出的本体概念集合进行匹配, 输入本体概念集合的匹配和输出本体概念集合的匹配在原理上基本上是一样的。假设, $\text{Request} = \{R_1, R_2, \dots, R_n\}$ 表示服务请求的输入本体概念集合, $\text{Service} = \{S_1, S_2, \dots, S_n\}$ 表示某个服务的输入本体概念集合, a 为相似度阈值。匹配算法描述如图 4 所示:

```

Inputs: Request, Service, a
Outputs: true or false
Procedure:
  Double sim = 1.0;
  For(int i=0; i<n; i++) {
    Double sim_0 = 0;
    For(int j=0; j<m; j++) {
      Double tmp = simFunction( Request, Service);
      If(tmp>sim_0) sim_0 = tmp;
    }
    Sim * = sim_0;
  }
  If(sim>=a) return true;
  Return false;

```

图4 输入输出匹配算法

其中, $\text{simFunction}(\text{Request}, \text{Service})$ 是(2)式的实现, 即本体概念集合相似度度量函数, 用于计算两个本体概念集合之间的相似度。

3 原型系统

原型系统采用三层架构:

(1) 用户接口层, 提供服务查询和服务注册的功能;

(2) 业务逻辑层;

(3) 数据存储层, 由 UDDI 及本体知识库组成。

其中业务逻辑层是框架的核心部分, 由 OWL-S/

UDDI 映射、服务请求解析、服务语义匹配、OWL-S 推理机和服务列表等 5 大模块组成, 各个模块的具体功能描述如下:

① OWL-S/UDDI 映射模块: 服务提供者提供的服务采用 OWL-S 描述, 而目前普遍采用的服务注册和发布系统是 UDDI, 可以对 OWL-S 的服务描述文档进行转换。将 OWL-S 文档中 `contactInformation`、`serviceName` 和 `textDescription` 分别映射到 UDDI `BusinessEntity` 的 `Contacts` 和 `BusinessService` 中的 `name` 和 `description`, 其它的语义信息同扩展的 UDDI `TModel` 类型中相应的元素建立映射关系, 这样就可以对服务进行注册;

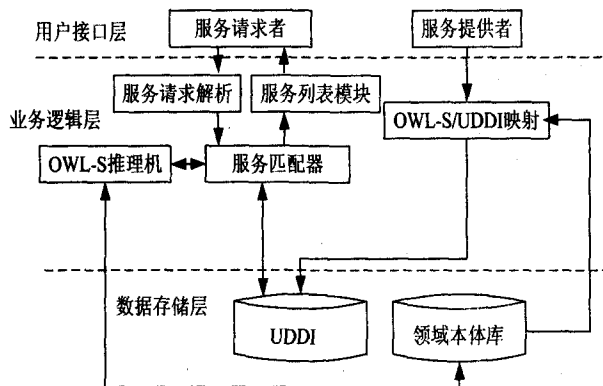


图5 基于OWL-S的Web服务发现系统原型系统框架

② 服务请求解析模块: 服务请求采用 OWL-S 文档进行描述, 解析并提取其中语义信息, 主要是领域本体概念以及其它一些基本信息, 也就是让计算机理解服务请求;

③ 服务匹配模块: 接收服务请求模块的带语义的服务请求, 先根据服务请求中基本信息, 比如服务类别、输入和输出参数的类型和数量等约束条件从 UDDI 中选出满足条件的候选服务, 这样可以提高效率, 避免对无关的服务进行匹配计算, 然后采用 OWL-S 推理机, 对服务请求和候选服务进行语义相似度计算的匹配, 将语义相似度满足阈值的 Web 服务传递给服务列表模块;

④ OWL-S 推理机: 结合领域本体知识库, 对给定的领域本体概念, 根据本体库中领域本体概念层次关系推导出它们之间的关系, 将推导结果传给匹配模块;

⑤ 服务列表模块: 将从服务匹配模块得到的所有满足请求阈值的候选服务, 按照相似度大到小进行排序, 展示给服务请求者。

4 实验结果

本实验在 Windows XP 环境下采用 Eclipse 开发工具、本体推理机 jena、服务注册中心 jUDDI 0.9rc3 和 XAMPP 服务器。从网络上搜集 100 个描述 Web 服务

的 WSDL 文件,服务包括天气查询、汽车价格查询、航班查询和图书查询等,用 OWL-S Editor 工具将 WSDL 文件转化成 OWL-S,再搜集 100 个 OWL-S 文件,将这 200 个服务进行注册到 jUDDI。然后分别基于传统关键字匹配的方法和文中方法进行 Web 服务发现实验,根据实验结果,对两种方法的查准率和查全率进行比较,结果如表 1 所示。实验结果表明文中的方法相比传统关键字匹配有较显著的优势。

表 1 实验结果

服务发现方法	查全率	查准率
关键字匹配	23.8%	28.3%
文中方法	83.3%	84.5%

5 结束语

传统的 Web 服务发现方法是基于简单的关键字匹配,服务查询的效率不高,难以满足用户的需求。文中提出了一种基于本体概念相似度计算的语义 Web 服务发现方法,采用 OWL-S 来描述服务请求和 Web 服务,运用本体概念相似度计算方法,使服务发现和匹配的精确度得到明显提高。实验证明了该方法的可行性和优越性。下一步的研究内容是进一步优化本体概念相似度匹配算法,提高语义 Web 服务发现的效率。

参考文献:

- [1] W3C. OWL-S: Semantic Markup for Web Services[EB/OL]. 2010-03-10. <http://www.w3.org/Submission/OWL-S/>.

(上接第 72 页)

算的发展有巨大的影响,促使人们对调度算法的深入研究。文中针对基本遗传算法,通过设定染色体的结构以及遗传算子的参数,在交叉操作时,增加了种群中染色体的多样性,并在变异操作之后,加入了针对特定问题的调整操作进行了精英选择,使算法具有了全局求解最优能力。

参考文献:

- [1] 徐志伟,冯百明,李伟. 网络计算技术[M]. 北京:电子工业出版社,2004:94-126.
- [2] 都志辉,李三立,陈瑜,等. 网络计算及其原型实现研究[J]. 计算机科学,2002,29(8):1-5.
- [3] 周建中,王树宗,石章松,等. 基于遗传算法的网络任务调度方法研究[J]. 舰船科学技术,2009,31(11):107-110.
- [4] Cao Junwei, Spooner D P, Jarvis S A, et al. Grid load balancing using intelligent agents[J]. Future Generation Computer Systems, 2005, 21(1): 135-149.
- [5] Pop F, Dobre C, Cristea V. Genetic algorithm for DAG scheduling in grid environments[C]//IEEE 5th International Confer-

- [2] 彭晖,陈立民,常亮,等. 基于动态描述逻辑的语义 Web 服务匹配研究[J]. 计算机研究与发展, 2008, 45(12): 2102-2109.
- [3] 邱田,李鹏飞,林品. 一个基于概念语义近似度的 Web 服务匹配算法[J]. 电子学报, 2009, 37(2): 429-432.
- [4] 陶望胜,陶先平,吕建. 基于规则与相似度的语义 web 服务匹配[J]. 南京大学学报, 2010, 46(2): 159-167.
- [5] 陈文字,张忠全,向涛,等. 基于相似度的语义 Web 服务发现技术研究[J]. 电子科技大学学报, 2010, 39(6): 896-899.
- [6] Kluseh M, Fries B, Syera K. OWLS-MX: A hybrid semantic Web service matchmaker for OWL-S services[J]. Journal of Web Semantics, 2009, 7(2): 121-133.
- [7] 张佩云,黄波,孙亚民. 一种基于语义与 Qos 感知的 Web 服务匹配机制[J]. 计算机研究与发展, 2010, 45(7): 780-787.
- [8] 李佳燕,史一民,李冠宇,等. 基于本体映射的语义 Web 服务发现框架[J]. 计算机工程与设计, 2009, 30(24): 5746-5750.
- [9] 张磊,夏士雄,周勇,等. 基于多本体的语义 Web 服务发现研究[J]. 计算机工程与应用, 2009, 45(11): 165-167.
- [10] 方常青. 基于语义的 Web 服务匹配和选择算法研究[D]. 长沙:中南大学, 2010.
- [11] 白东伟. 基于语义的 Web 服务匹配与发现技术研究[D]. 北京:北京邮电大学, 2007.
- [12] 潘善亮,张迎新,俞晓峰. 基于 OWL-S 的 Web 服务匹配研究[J]. 计算机工程, 2010, 36(12): 48-51.

ence on Intelligent Computer Communication and Processing. [s.l.]: [s.n.], 2009.

- [6] 王猛. 基于改进遗传算法的网格资源调度策略算法研究[J]. 电脑与电信, 2010, 5(1): 35-37.
- [7] Zhang Yi, Zhang Jun, Shu Xingzu, et al. Optimization of Intelligent Compactness Control Rule of Vibratory Roller Based on Genetic Algorithm Method[C]//Proceedings of the 2009 Fifth International Joint Conference on INC, IMS and IDC. [s.l.]: [s.n.], 2009.
- [8] 张阳,黄文明,兰静. 一种基于改进遗传算法的网格任务调度策略[J]. 计算机系统应用, 2009(7): 37-40.
- [9] 叶春晓,陆杰. 基于改进遗传算法的网格任务调度研究[J]. 计算机科学, 2010, 37(7): 233-235.
- [10] 刘海迪,杨裔,马生峰,等. 基于分层遗传算法的网格任务调度策略[J]. 计算机研究与发展, 2008, 45(sup): 35-39.
- [11] 杨琴,王海瑞. 基于遗传算法的网格任务调度研究[J]. 山西电子技术, 2009(1): 95-96.
- [12] 韩瑞峰. 遗传算法的原理与应用实例[M]. 北京:兵器工业出版社, 2010.