

基于三次匹配的语义 Web 服务发现模型

高 洋, 黄映辉

(大连海事大学 信息科学技术学院, 辽宁 大连 116026)

摘 要:针对传统 Web 服务发现机制 UDDI 存在之不足, 提出一种基于 OWL-S 的 Web 服务发现模型, 其特点是对请求服务与发布服务的功能属性、基本描述、质量属性依次进行匹配。在服务功能属性的匹配中, 通过引入关系数据库, 将服务功能属性的匹配转换为关系数据库的简单查询与选择操作, 节省了服务匹配时间; 在服务基本描述的匹配中, 通过过滤与请求服务无关的服务, 缩小了服务匹配范围; 在服务质量属性的匹配中, 通过分析服务的主要 QoS 属性, 可使服务请求者在服务质量不同的待选服务中发现所期待的服务。

关键词:服务发现; 服务匹配; 服务功能属性; 服务基本描述; QoS 属性

中图分类号: TP393

文献标识码: A

文章编号: 1673-629X(2009)09-0122-03

A Model for Semantic Web Service Discovery Based on Three Matching

GAO Yang, HUANG Ying-hui

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

Abstract: Be directed against the defects of UDDI in traditional Web service discovery, a model for Web service discovery based on OWL-S is proposed. The feature of the model is that the requested service and the published services achieve the matching of functional attributes, the matching of basic description, and the matching of QoS in turn. In the matching of functional attributes, due to the introduction of a relational database, the functional attributes matching of service will be converted to the simple querying and selecting operations in the relational database, and the time of matching could be saved; In the matching of service basic description, through filtering the services which have nothing to do with the requested service, the scope of services matching would be narrowed; In the matching of QoS, with analyzing the main QoS attributes of the services, the service requestors could find out the service needed in the offered services with different qualities.

Key words: service discovery; service matching; service functional attribute; service basic description; QoS attribute

0 引 言

随着 Web 服务应用的发展, Web 服务发现已成为研究的热点。由于现有成熟的 Web 服务发现机制缺乏语义, 目前的研究大多集中在语义 Web 服务发现上。M. Paolucci 等提出了一个语义 Web 服务发现模型^[1], 但随着注册到 UDDI 中的 Web 服务不断增加, 处理每个查询的时间也会延长; 同时该模型在进行服务匹配时缺乏服务质量属性的匹配。文中提出了一个修正模型, 其特征是通过实现服务的三次匹配, 即服务

功能属性的匹配、服务基本描述的匹配、服务质量(QoS)属性的匹配, 不仅实现了服务的语义匹配, 而且可使服务请求者在服务质量不同的待选服务中发现所期待的服务。

1 模型的提出

描述语义 Web 服务的规范语言主要有 WSDL-S、OWL-S 和 WSMO。目前应用最多的 OWL-S 是基于 OWL 语言的 Web 服务本体, 可将一个 Web 服务描述成 Service Profile、Service Model、Service Grounding 三个视图, 其中 Service Profile 用于支持 Web 服务的发现, 即主要描述服务能做什么, 具体为三个方面: 服务基本描述、服务功能属性和服务非功能属性。在非功能属性中用可扩展属性 serviceParameter 包含服务质量(QoS)属性^[2]。文中的研究思路是针对 Service Profile 所描述的三类属性依次进行匹配, 即服务功能属性

收稿日期: 2009-01-18; 修回日期: 2009-03-23

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

作者简介: 高 洋(1984-), 女(回族), 内蒙古宁城人, 硕士研究生, 研究方向为智能信息处理; 黄映辉, 教授, 研究方向为智能信息处理。

的匹配、服务基本描述的匹配和服务质量属性的匹配。所提出的语义 Web 服务发现模型如图 1 所示。

该模型主要由通信模块、OWL-S/UDDI 转换器、UDDI 注册中心、本体、三个匹配模块(功能属性匹配模块、基本描述匹配模块、QoS 属性匹配模块)组成。如图 1 所示,由本体得到概念关系树,然后将其存储到关系数据库中^[3]。由 OWL-S 标记的服务描述通过通信模块判断是服务发布描述(description of service published)还是服务请求描述(description of service requested)。若是前者,首先让其进入功能属性匹配模块更新关系数据库中的概念参数,然后通过 OWL-S/UDDI 转换器实现 Service Profile 中的元素与 UDDI 信息结构中的相应元素的一对一映射以便将服务注册到 UDDI 中^[4]。若是后者,首先让其进入功能属性匹配模块,将大于或等于规定的最小匹配度的服务传输到基本描述匹配模块进行匹配,然后将筛选后的服务送入到 QoS 属性匹配模块进行匹配,最后将满足服务请求者定义的最小匹配度的服务及其匹配度返回给服务请求者。

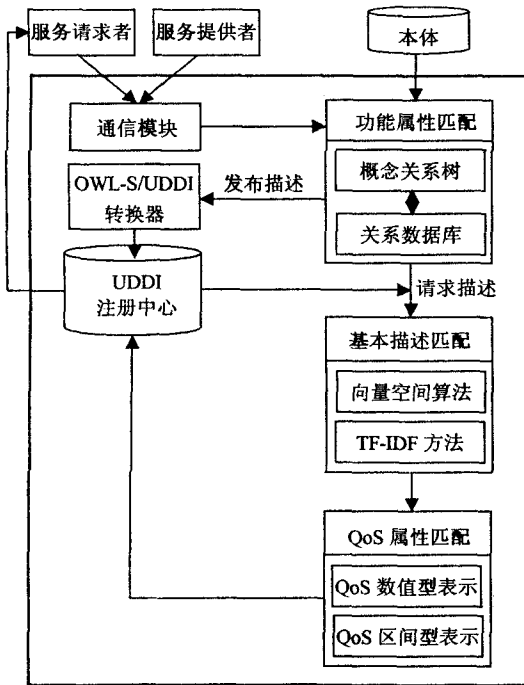


图 1 基于三次匹配的语义 Web 服务发现模型

2 三次匹配及其过程

2.1 三次匹配

服务发现的过程主要是进行服务匹配。现有 Web 服务发现机制缺乏语义,服务匹配是基于关键字的,服务请求者常常无法找到所期待的服务。即使在 M. Paolucci 的语义 Web 服务发现的模型^[2]中,虽然实现

了 Web 服务的语义匹配,但在匹配过程中,服务请求者要与 UDDI 注册中心中的所有服务进行遍历匹配,其中包括一些与服务请求者所需服务并无关联服务,这就大大增加了服务匹配的时间。此外,对于提供相同功能的服务其服务质量也是各异的,所以为使服务请求者找到其所期待的服务,还有必要对服务质量进行匹配。

本模型所提出的三次匹配是指:首先,进行服务功能属性的匹配,在此匹配模块中,通过引入关系数据库,将服务功能属性的匹配转换为关系数据库的简单查询与选择操作,节省了服务匹配时间;其次,进行服务基本描述的匹配,其作用是对服务进行进一步的筛选,缩小了待选服务的范围;最后,进行服务的 QoS 属性匹配,通过对服务的主要 QoS 属性进行匹配,使服务请求者找到适合其需求的服务。经过三次匹配,将所获得的待选服务及其与服务请求者描述服务的匹配度返回给服务请求者供其选择。

2.2 服务功能属性的匹配

在服务功能属性匹配过程中,必须满足两个条件:服务请求者的输入可以满足服务发布者的输入,服务提供者的输出可以满足服务请求者的输出^[3]。所以,要进行服务输入和服务输出两个功能属性的匹配。这种匹配采用的是前述 M. Paolucci 的语义匹配思想^[1],服务请求者与服务提供者之间的功能属性(输入或输出)的匹配程度取决于描述它们的概念之间的匹配度,并且认为概念之间有四种匹配度。

现以服务输出属性为例来说明服务功能属性匹配的过程。设 R_o 代表服务请求者的一个输出参数, P_o 代表服务发布者的一个输出参数。精确匹配(exact): R_o 是 P_o 的等价概念,或者 R_o 是 P_o 的直接子概念;包含匹配(plugin): R_o 是 P_o 的非直接子概念;插入匹配(subsume): R_o 是 P_o 的父概念;失败匹配(fail): R_o 与 P_o 无关。

由于服务发布对实时性要求不高,所以提出在服务发布阶段将相应本体映射到关系数据库中^[5],对关系数据库中的每一个元素添加两个表 Input_node_information 和 Output_node_information。在 Input_node_information 表中用二元组 $\langle S, \text{匹配度}_i \rangle$ 表示概念做为服务提供者发布服务 S 的输入参数的匹配度为匹配度_i;在 Output_node_information 表中用二元组 $\langle S, \text{匹配度}_o \rangle$ 表示概念作为服务提供者发布服务 S 的输出参数的匹配度为匹配度_o^[3]。此处定义精确匹配的匹配度为 1.0、包含匹配的匹配度为 0.5、插入匹配的匹配度为 0.25、失败匹配的匹配度为 0。并且规定二元组中的匹配度初始值为 0。

图 2 是宠物交易领域本体的一个片段。设某服务提供者提供的服务 S 是销售宠物服务, 它的输出是“宠物”和“价格”。于是在关系数据库中查找“宠物”这个概念, 将二元组 $\langle S, 1.0 \rangle$ 添加到“宠物”这个概念的 Output_node_information 表中, 这表明“宠物”概念做为服务 S 的输出是精确匹配。同理更新“价格”概念的 Output_node_information 表。根据概念之间的四个匹配度以及概念之间的关系, 更新关系数据库中其它概念与服务 S 输出参数之间的匹配度。由图 2 可见“动物”概念是“宠物”概念的父概念, 则将二元组 $\langle S, 0.25 \rangle$ 添加到“动物”概念的 Output_node_information 表中。以此类推。这样, 关系数据库中的每个概念作为服务 S 的输入和输出都有一个匹配值。

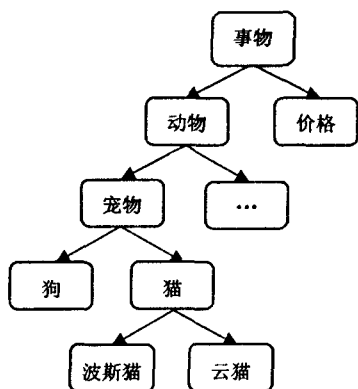


图 2 宠物交易领域本体的片段

由于要求服务提供者的输出可以满足服务请求者的输出, 则描述服务请求者输出属性的概念至少是描述服务提供者输出属性的概念的子概念。所以, 描述服务请求者输出属性的概念与描述服务提供者输出属性的概念之间的匹配至少为插入匹配, 即二者匹配度不能低于 0.25。通过以上分析, 在服务查找时, 首先从关系数据库中描述服务请求者输出属性的所有概念的 Output_node_information 表, 然后在这几个表中取共有服务, 同时剔除匹配度小于 0.25 的服务, 得到服务集 AV_1 。同理, 对描述服务请求者输入属性的所有概念的 Input_node_information 表进行类似的操作, 得到服务集 AV_2 。最后取 AV_1 与 AV_2 的交集 AV , 将其中的服务输送到服务基本描述的匹配中。

2.3 服务基本描述的匹配

在服务基本描述的匹配过程中, 选取服务分类(serviceCategory)、服务名称(serviceName)与文本描述(textDescription)三个主要属性进行匹配。由于它们均用字符串表示, 所以匹配过程实质上是进行文本相似性比较。匹配采用向量空间算法^[6]: 文本 $D_p \{w_{p1}, w_{p2}, \dots, w_{pm}\}$ 与文本 $D_q \{w_{q1}, w_{q2}, \dots, w_{qn}\}$ 的相似度用两个向量的夹角余弦表示为:

$$\text{Sim}(D_p, D_q) = \frac{\sum_{i=1}^n w_{pi} * w_{qi}}{\sqrt{\sum_{i=1}^n w_{pi}^2} * \sqrt{\sum_{i=1}^n w_{qi}^2}}$$

其中 w_{pi} 表示文本 D_p 的第 i 个特征项的权重。可见, $\text{Sim}(D_p, D_q)$ 越大, 文本 D_p 与 D_q 相似程度越大。此处用 TF-IDF 方法^[4]提取文本的特征项并计算特征项对应的权重。这样分别计算服务提供者和服务请求者的服务分类、服务名称、文本描述三个匹配度, 然后取三者的综合匹配度。将综合匹配度满足服务请求者定义的最小匹配度的服务输送到服务的 QoS 属性匹配中。

2.4 服务 QoS 属性的匹配

不同的服务请求者对服务质量(QoS)的要求是有差异的。为满足服务请求者的需求, 应进行服务质量属性的匹配。选择主要的 QoS 属性即时间(time)、费用(cost)、可靠性(reliability)^[7]进行匹配。不同的服务请求者对这些属性有不同的表达方式, 既可用数值型表示(例如 $\text{cost} = 200$ 元), 也可用区间型表示(例如 $\text{time} \in [100, 200]$)。故针对 QoS 属性两种不同的表达方式, 采用不同的相似性度量方法。

对于数值型的 QoS 属性, 可用 MAE(mean absolute error)算法^[7,8]来获得服务提供者的 QoS 属性 PQ 与服务请求者的 QoS 属性 RQ 之间的相似度, 即

$$d = 1 - \frac{|PQ - RQ|}{RQ}$$

可见, d 值越大表示 PQ 与 RQ 的相似度越大。

对于区间型的 QoS 属性, 可用属性区间数之间的相离度^[9]来表明两个区间是否相似。设两个区间 $a = [a_l, a_u]$ 、 $b = [b_l, b_u]$, 则二者的相离度为

$$D(a, b) = \sqrt{(a_l - b_l)^2 + (a_u - b_u)^2}$$

显然 $D(a, b)$ 越小, 说明区间 $[a_l, a_u]$ 与区间 $[b_l, b_u]$ 之间相似程度越大。

3 结束语

在基于三次匹配的语义 Web 服务发现模型中, 应用 OWL-S 标记 Web 服务, 实现 Web 服务的语义描述; 应用通信模块判断服务描述, 区分为服务发布描述与服务请求描述; 应用 OWL-S/UDDI 转化器进行 Service Profile 中元素与 UDDI 的映射, 实现 UDDI 的语义扩展; 应用三次匹配, 实现服务的功能属性、基本描述、质量属性的匹配。在服务功能属性的匹配过程中, 由于基于语义扩展的 UDDI 中注册了大量的服务, 在服务进行功能属性的匹配时会浪费大量时间, 故引

(下转第 194 页)

- 没有该账号或者密码错误

begin

Select @ReturnType = 'Nook'

Select @ReturnDes = '没有该账号或者密码错误,请与管理
员联系!'

Select @ReturnType

Return

end

GO

4.2 用户结账存储过程设计

在用户结账存储过程中,有 1 个入口参数,两个出口参数,该存储过程算法较为简单,就是根据账号上机登录信息计算上机实际消费金额,在计算中要考虑不同机房、不同机器、不同账号要有不同的收费标准,将计算好的消费金额处理用户信息表和用户登录表的信息。在结账中要处理停电情况和学生机和服务器长时间没有通信的情况。

4.3 用户刷新存储过程设计

通过用户刷新可以计算该用户的消费金额,并根据所消费金额计算剩余金额,如果用户剩余金额不足 0.2 元提示“用户金额不足”信息给用户,如果剩余金额低于 0.02 元时用户结账关机。

5 结束语

在机房管理中除了存储过程实现机房管理的业务逻辑^[9],还要设计学生机的身份验证机制,在学生机启动后不能绕过身份验证就使用计算机,学生机在使用系统后要防止用户关闭客户端程序,并且在服务器和客户端同时计费,防止用户登录后与服务器断开连接后盗用机时。本系统的功能已经实现,并且本系统已

经在我校计算中心运行一年多,效果良好。因此可以得出以下几点结论:

- 1) 采用存储过程后,存储过程只向应用程序提供接口,修改业务逻辑较为方便;
- 2) 修改业务逻辑不用修改应用程序,方便程序升级;
- 3) 采用存储过程可以很好地控制系统并发性;
- 4) 采用存储过程业务处理效率高。

参考文献:

- [1] Deja. SQL Server2000 高级编程技术[M]. 北京:清华大学出版社,2002:102-108.
- [2] 汪维富,黄海于,陈娟,等.基于存储过程的高性能数据库应用模型研究[J]. 计算机工程与设计,2008,29(10):2573-2575.
- [3] 何健,王家华.树型层次结构数据中遍历子树结点的方法[J]. 计算机技术与发展,2008,18(4):95-97.
- [4] 夏义全.数据库应用系统优化方法的研究[J]. 计算机技术与发展,2008,18(7):149-152.
- [5] 孙伟东,马宗民.利用存储过程实现简单分布式计算的方法[J]. 沈阳航空工业学院学报,2008(1):53-55.
- [6] 苗专生.学校机房管理系统的设计与实现[J]. 苏州科技学院学报:自然科学版,2006(2):77-80.
- [7] Funderburk J E, Malaika S, Reinwald B. XML programming with SQL/XML and Xquery[J]. IBM Systems Journal,2002,41(4):642-666.
- [8] Beauchemin B, Berglund N, Sullivan D. A First Look at SQL Server 2005 for Developers[M]. Boston: Addison - Wesley, 2004.
- [9] Blaha M. Designing and Implementing SoftcodedValues[M]. [s.l]:[s.n.],2006.

(上接第 124 页)

入关系数据库,使服务功能属性的匹配转换成关系数据库的简单查询与选择操作,在此匹配模块中不但实现了语义匹配,而且在很大程度上缩短了查找时间;服务基本描述的匹配的作用是过滤,剔除与服务请求无关的服务;服务质量属性的匹配充分考虑了服务请求者的不同需求,可以提高查准率。

参考文献:

- [1] Srinivasan N, Paolucci M, Sycara K. Semantic Web service discovery in the OWL-S IDE[EB/OL]. [2005-10-04]. <http://iswc2004.semanticweb.org/demos/10/paper.pdf>.
- [2] 贾晖.扩展 OWL-S 的 Web 服务质量属性描述方法研究[J]. 现代电子技术,2007,30(16):541-542.
- [3] 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,2007,9(7):2103-2107.

- [4] 罗欣,夏德麟.基于词频差异的特征选取及改进的 TF-IDF 公式[J]. 计算机应用,2005,25(9):2301-2303.
- [5] 陶皖,姚红燕.OWL 本体关系关系数据库存储模式设计[J]. 计算机技术与发展,2007,17(2):111-114.
- [6] 李春梅.基于本体的语义 Web 服务发现及选取的研究[D]. 桂林:广西师范大学,2006.
- [7] Zeng I, Benat B. QoS-Aware Middle-ware for Web Service Composition[J]. IEEE Transaction on Software Engineering,2004,30(5):311-327.
- [8] 吴芸.基于语义的 Web 服务发现研究[D]. 西安:西北大学,2007.
- [9] 蓝继斌,刘芳.区间数可能度的二维定义[J]. 数学的实践与认识,2007,27(24):67-70.