

## 基于领域本体的Web服务发现

艾未华<sup>1,2</sup>, 宋自林<sup>2</sup>, 魏磊<sup>2</sup>, 吴量<sup>2</sup>

(1. 解放军理工大学气象学院 南京 211101; 2. 解放军理工大学指挥自动化学院 南京 210007)

**【摘要】** Web服务应用的一个关键问题是如何发现需要的服务。传统的基于关键字的服务发现机制UDDI缺乏语义支持, 搜索效率低。据此该文研究基于领域本体的Web服务发现技术, 对发布服务和请求服务都采用OWL-S描述以支持基于语义的服务搜索能力, 并且提出了一种基于图匹配的Web服务发现算法。试验结果表明服务发现方法在查准率和查全率方面都要优于UDDI。

**关键词** 二部图匹配; 领域本体; 服务发现; Web服务  
**中图分类号** TP311 **文献标识码** A

## Web Service Discovery Based on Domain Ontology

AI Wei-hua<sup>1,2</sup>, SONG Zi-lin<sup>2</sup>, WEI Lei<sup>2</sup>, WU Liang<sup>2</sup>

(1. Institute of Meteorology, PLA University of Science and Technology Nanjing 211101;  
2. Institute of Command Automation, PLA University of Science and Technology Nanjing 210007)

**Abstract** The important step towards Web service implement is the discovery of services. Traditional keywords-based Web service discovery mechanism, Universal Description, Discovery, and Integration(UDDI), lacks semantic descriptions and has low searching efficiency. We study the Web service discovery based on domain ontology. The advertisement services and requested services are described by (Ontology Web Language for Services, OWL-S) OWL-S to support semantic searching. A new web services discovery algorithm based on graph matching is presented. The experiments show that the proposed algorithm can achieve better results than UDDI in precision and recall.

**Key words** bipartite graph matching; domain ontology; service discovery; Web service

随着Internet的飞速发展, Web应用逐步由集中式向分布式转变。Web服务<sup>[1]</sup>是一种崭新的分布式计算模型, 是自包含的、模块化的应用程序, 它可以在Web中被描述、发布、查找以及调用, 执行特定的任务, 遵守具体的技术规范, 这些规范使得Web服务能与其他兼容的应用程序进行互操作。Web服务发现是Web服务系统架构中的一个重要部分。统一描述、发现和集成规范(Universal Description, Discovery and Integration, UDDI<sup>[2]</sup>)是一种基于关键词和简单分类的服务发现机制, 它通过对代理上的服务注册信息进行关键词精确匹配实现服务发现。虽然基于UDDI的服务查找方法使用非常方便, 但是却存在查准率和查全率不高的问题。主要原因是该方法缺乏对服务的精确描述, 不能比较服务之间的相似或符合程度。

语义Web<sup>[3]</sup>是对现有Web的扩展, 它与传统Web

的不同, 在于信息语义在语义Web环境下能够很好地加以定义, 并使得计算机之间能够更好地协同工作, 其目标是让Web上的信息能够被机器理解, 从而实现Web信息的自动处理。本体是语义Web中的重要概念, 是描述语义Web中语义知识的建模手段, 它形式化定义了领域内共同认可的知识, 是语义Web体系中的核心。将语义Web中基于本体的知识标识手段来描述Web服务的语义, 使计算机可以理解Web服务, 就可克服当前Web服务中存在的问题, 从而支持服务的自动发现, 提高Web服务的查询效率。文献[4]提出用DAML-S语言描述服务, 通过计算服务的属性和接口的输入输出概念匹配程度得到服务匹配结果。该方法定义了四种简单的概念匹配程度, 对概念之间的关系描述局限于概念之间的包含关系。

本文研究基于领域本体的Web服务发现技术,

收稿日期: 2006-11-18

基金项目: 总装备部武器装备部预研基金资助项目

作者简介: 艾未华(1979-), 男, 博士生, 讲师, 主要从事语义Web、数据库技术等方面的研究。

用Web服务本体描述服务的Web本体语言(Ontology Web Language for Services, OWL-S<sup>[5]</sup>)和领域本体描述服务。采用概念的语义相似度量服务属性和接口输入输出概念之间的相符程度,并在此基础上引入服务二部图的概念,将两个服务的匹配转化成二部图的匹配。

### 1 基本概念

定义 1 领域本体是对特定领域内概念及概念间关系的精确描述,可以用五元组表示:  $O = \{C, R, H^c, rel, A^o\}$ 。其中  $C$  表示概念的集合;  $R$  表示关系的集合;  $H^c$  表示概念层次;  $rel$  表示概念间的关系;  $A^o$  表示本体公理。服务属性和接口输入输出等都是用领域本体来描述。

定义 2 服务发现首先需要一种语言描述服务。本文用OWL-S描述服务,并定义 Profile-A 和 Profile-R 分别为发布服务和请求服务的Profile。服务的语义匹配就是分析 Profile-A 和 Profile-R 语义描述,判断它们之间的相符和相似程度。

定义 3 一个服务Profile包含多组描述服务属性和接口参数的概念,如输出参数组和输入参数组等。用  $A(a_1, a_2, \dots, a_m)$  表示发布服务的 Profile-A 的输出概念的集合;  $R(r_1, r_2, \dots, r_n)$  表示请求服务的 Profile-R 的输出概念的集合。定义服务参数对相似度 (Similarity of Parameter Pairing, SPP),  $S_{SPP_{i,j}}(a_i, r_j) = S_{Sim}(a_i, r_j)$ , 其中  $a_i$  为  $A(a_1, a_2, \dots, a_m)$  中的一个概念;  $r_j$  为  $R(r_1, r_2, \dots, r_n)$  中的一个概念;  $S_{Sim}(a_i, r_j)$  是实现概念  $a_i$  和  $r_j$  之间语义相似度的度量。服务相似度计算的基础是概念间语义相似度。本文给出基于领域本体的概念间语义相似度公式:  $S_{SPP_{i,j}}(a_i, r_j) = a/(d+a)$ 。其中  $S_{SPP_{i,j}}(a_i, r_j)$  表示两个概念  $a_i$  和  $r_j$  的语义相似度;  $d$  是两个概念之间的语义距离;  $a$  是一个可以调节的参数;  $0 < S_{SPP_{i,j}}(a_i, r_j) < 1$  且  $S_{SPP_{i,j}}(a_i, r_j) = S_{SPP_{j,i}}(a_j, r_i)$ 。如果两个概念的语义相等,那么其相似度为1;如果两个概念不满足语义包含或相交关系,则其相似度为0。用概念相似度矩阵表示  $A(a_1, a_2, \dots, a_m)$ 、 $R(r_1, r_2, \dots, r_n)$  之间的概念相似度的二维关系,服务参数概念相似度矩阵为:

$$\begin{bmatrix} S_{SPP_{1,1}} & r_1 & r_2 & \dots & r_n \\ a_1 & S_{SPP_{11}} & S_{SPP_{12}} & \dots & S_{SPP_{1n}} \\ a_2 & S_{SPP_{21}} & S_{SPP_{22}} & \dots & S_{SPP_{2n}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_m & S_{SPP_{m1}} & S_{SPP_{m2}} & \dots & S_{SPP_{mn}} \end{bmatrix}$$

定义 4 计算两个服务输出参数匹配程度的实质转变成计算  $A(a_1, a_2, \dots, a_m)$  和  $R(r_1, r_2, \dots, r_n)$  两组概念之间的相似程度,而这个相似度计算可用加权二部图的最优匹配方法解决,图1所示为二部图模型。定义  $G_{m,n} = (A_m, R_n, E)$  为二部图,其中  $A_m = (a_1, a_2, \dots, a_m)$ ;  $R_n = (r_1, r_2, \dots, r_n)$ ;  $E = (e_{11}, e_{12}, \dots, e_{m,n})$ 。  $A_m$  和  $R_n$  图顶点的集合;  $E$  是连接任意两个概念顶点的边的集合,且任意边  $e_{ij}$  的二部图权值为  $S_{SPP_{i,j}}(a_i, r_j)$ , 即是任意两个概念  $a_i$  和  $r_j$  之间的语义相似度。

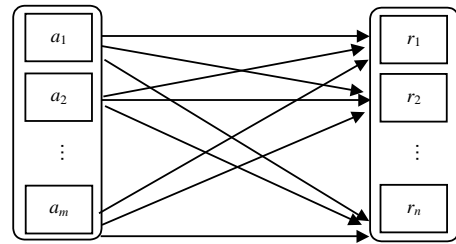


图1 二部图模型

### 2 服务匹配算法

#### 2.1 最大匹配和最佳匹配

最大匹配和最佳匹配是二分图中的两个经典问题。所谓匹配就是指给定一个二分图  $G_{m,n} = (A_m, R_n, E)$ , 在  $G_{m,n}$  的一个子图中,  $M$  的边集  $E$  中的任意两条边都不依附于同一个顶点, 则称  $M$  是一个匹配。选择这样的边数最大的子集称为图的最大匹配问题, 如图2所示。如果边上带权, 找出权和最大的匹配叫做求最佳匹配, 如图3所示。

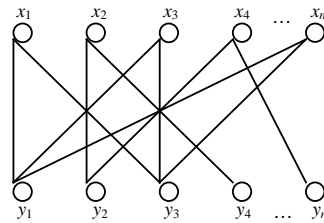


图2 最大匹配

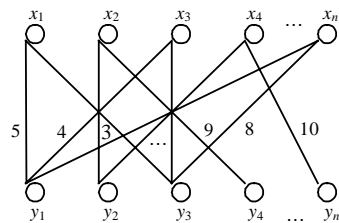


图3 最佳匹配

#### 2.2 概念组相似度算法

由以上分析不难看出, 计算两组概念的相似程度并求其相似度的最大值的问题类似于二部图的最

优匹配问题。最优匹配方法的实质是在一一对应的前提下，衡量两个概念组所能达到的最大相似度。本文将二部图的思想应用到Web服务匹配中，将两个服务相似程度的度量转化成多个概念组的匹配程度的计算。Kuhn-Munkres<sup>[6]</sup>算法是计算加权二部图的最优匹配的经典方法，本文将该算法用于计算服务的两组概念的相似程度。由定义4知： $G_{m,n} = (A_m, R_n, E)$ 为一个加权二部图， $A_m$ 和 $R_n$ 图顶点的集合， $E$ 是边的集合，其中 $A_m = (a_1, a_2, \dots, a_m)$ ， $R_n = (r_1, r_2, \dots, r_n)$ ； $S_{SPP_{i,j}}(a_i, r_j)$ 是任意边的权值。具体算法步骤为：

(1) 给出初始标号  $l(a_i) = \max_j S_{SPP_{i,j}}(a_i, r_j), l(r_j) = 0, i, j = 1, 2, \dots, t, t = \max(n, m)$ ；(2) 用Hungarian算法<sup>[6]</sup>求边集  $E_l = \{(a_i, r_j) | l(a_i) + l(r_j) = S_{SPP_{i,j}}(a_i, r_j)\}$ ， $G_l = (A_m, R_n, E_l)$ 及 $G_l$ 中的完备匹配 $M$ ；(3) 若 $M$ 是 $G_l$ 中的完备匹配，则 $M$ 即是 $G$ 的最佳匹配，计算结束，否则进行下一步；(4) 在 $A_m$ 中找 $M$ 的非饱和点 $a_0$ ，令 $A \leftarrow \{a_0\}, B \leftarrow \emptyset$ ， $A, B$ 是两个集合；(5) 若 $P_{G_l}(A) = B$ ，则转步骤(9)，否则进行下一步，其中 $P_{G_l}(A) \subseteq R_n$ 是与 $A$ 中结点邻接的结点集合；(6) 找一结点 $r \in P_{G_l}(A) - B$ ；(7) 若 $r$ 是 $M$ 饱和点，则找出 $r$ 的配对点 $z$ ，令 $A \leftarrow A \cup \{z\}, B \leftarrow B \cup \{r\}$ ，转步骤(5)，否则进行下一步；(8) 存在一条从 $a_0$ 到 $r$ 的可增广路径 $R$ ，令 $M \leftarrow M \oplus E(R)$ ，转步骤(3)；(9) 按下式计算 $\alpha$ 值： $\alpha = \min_{\substack{a_i \in A \\ r_j \in N_{G_l}(A)}} \{l(a_i) + l(r_j) - S_{SPP_{i,j}}(a_i, r_j)\}$ ，修改

标号： $l'(v) = \begin{cases} l(v) - \alpha & \text{若 } v \in A \\ l(v) + \alpha & \text{若 } v \in B \\ l(v) & \text{其他} \end{cases}$ 。根据 $l'$ 求 $E_{l'}$ 及

$G_{l'}$ ；(10)  $l \leftarrow l', G_l \leftarrow G_{l'}$ ，转步骤(6)。

### 2.3 算法的改进

Kuhn-Munkres算法从理论上能很好地解决两组概念的相似度的度量问题，但是在实际应用该算法时还需对其进行相应的改进。本文从三个方面进行了算法的改进和调整：(1) Kuhn-Munkres算法的输入必须是一个完全二部图，即在二部图 $G_{m,n} = (A_m, R_n, E)$ 中 $m$ 的值必须等 $n$ ，但是在服务匹配过程中，请求服务的参数个数与发布服务的参数个数通常是不等的。假设 $n = \max(m, n)$ ，令 $A_n = (a_1, a_2, \dots, a_m, a_{m+1}, \dots, a_n)$ ， $R_n = (r_1, r_2, \dots, r_n)$ 。赋值 $S_{SPP_{i,j}}(a_i, r_j) = 0$ ，其中 $i = m+1, m+2, \dots, n$ ；(2) 概念相似度 $S_{SPP_{i,j}}(a_i, r_j)$ 的值表示概念对 $a_i$ 和 $r_j$ 的相关程度。在进行服务匹配时，如果某个概念对的语义相似度 $S_{SPP_{i,j}}(a_i, r_j)$

低于域值 $t$ ，则令其值为零，即 $S_{SPP_{i,j}}(a_i, r_j) = 0$ ；

(3) Kuhn-Munkre算法利用Hungarian算法计算边集 $E_l = \{(a_i, r_j) | l(a_i) + l(r_j) = S_{SPP_{i,j}}(a_i, r_j)\}$ ， $G_l = (A_m, R_n, E_l)$ 及 $G_l$ 中的完备匹配 $M$ 。如果计算出的完备匹配 $M$ 中出现边的权值为零的情况，则将该完备匹配 $M$ 舍弃，继续采用Hungarian算法计算完备匹配，直至得到边的权值不为零的完备匹配或返回匹配失败。求出 $G_{m,n} = (A_m, R_n, E)$ 的最优匹配 $M$ 后，把 $M$ 每条边的权值 $S_{SPP_{i,j}}(a_i, r_j)$ 相加，可以求得 $G_{m,n} = (A_m, R_n, E)$ 的最大权值。假设 $S_{\text{output}}$ 是服务输出的两个概念组匹配最大权值， $S_{\text{input}}$ 是服务输入的两个概念组匹配的最大权值，且 $S_{\text{output}}$ 和 $S_{\text{input}}$ 都已归一化，那么请求服务和发布服务的总体相似度是 $S_{\text{total}} = \alpha S_{\text{output}} + \beta S_{\text{input}}$ ，其中 $\alpha + \beta = 1, 0 < \alpha < 1, 0 < \beta < 1$ 。

## 3 系统框架及试验结果

本文Web服务匹配实现的总体思路是：当一个服务请求提交之后，首先将请求服务的输出和已发布服务的输出进行基于概念相似度计算的匹配；在找到输出合适的已发布服务之后，再进行请求服务的输入和该发布服务的输入的比较，最后得到匹配结果。图4所示是基于语义相似度的服务发现的实现框架。服务请求者和发布服务都用OWL-S作为描述语言，对服务的输入和输出等参数用领域本体进行描述。服务解析模块用Jena API<sup>[7]</sup>加载请求服务和发布服务，并解析出输入和输出等描述服务的参数信息。相似度计算模块计算任意两个概念之间的语义相似度，得到服务参数概念相似度矩阵。服务匹配模块利用Kuhn-Munkres算法计算请求服务和发布服务之间的匹配程度，最后得到输出结果。

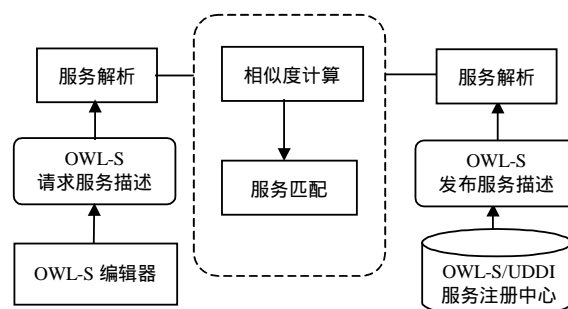


图4 基于本体的服务发现框架

本文实现了所提出的基于领域本体的Web服务发现算法，并与传统的基于关键字的Web服务发现策略UDDI进行比较。用查准率和查全率作为度量

Web服务发现性能的指标。查全率是指查询返回符合查询条件的Web服务与查询返回Web服务总数量的比率,查准率是指查询返回符合查询条件的Web服务与测试样本集中符合查询条件的Web服务的比率。分别用本文提出的算法和基于关键字的匹配的Web服务发现方法UDDI进行测试,得到的平均查全率为89%和56%,平均查准率为76%和43%,如图5所示。试验结果表明,本文提出的服务发现方法在查准率和查全率方面都要优于UDDI。

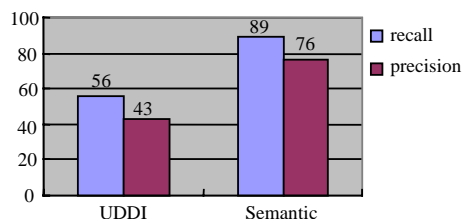


图5 性能测试查准率和召回率比较图

## 4 结束语

本文介绍了Web服务的本体语义描述和基于语义相似度的服务发现匹配算法。传统的基于WSDL的服务描述语言缺乏对服务的语义描述,本文用语义本体来描述服务以解决该问题,基于本体的语义匹配算法可以实现传统的字符串匹配算法不具备的精确性和灵活性。本文提出的算法在查准率和查

全率上要优于现有UDDI上的基于关键字的服务发现机制。下一步研究还将提供更加详细的服务匹配信息以提高服务匹配的精确度以及研究在语义匹配的基础上进行服务的自动组合和执行等。

## 参 考 文 献

- [1] HAAS H. Web services activity statement[EB/OL]. <http://www.w3.org/2002/ws/Activity>, 2002-04-12.
- [2] UDDI org. Universal description, discovery, and integration of businesses for the web[EB/OL]. <http://www.uddi.org/>, 2004-10-01.
- [3] LEE T B, HENDLER J, LASSILA O. The semantic Web[J]. *Scientific American*, 2001, 284(5): 34-43.
- [4] PAOLUCCI M, KAWAMURA T, Payne T R, et al. Semantic matching of Web services capabilities[C]//In: Proceedings of the 1st International Semantic Web Conference (ISWC). Sardinia, Italia: [s. n.], 2002: 34-43.
- [5] OWL Services Coalition. OWL-S: Semantic markup for Web services[EB/OL]. <http://www.daml.org/services/owl-s/1.0/>, 2003-03-15.
- [6] DAVES A M. The optimal assignment problem[EB/OL]. <http://www.math.uwo.ca/~mdawes/courses/344/kuhn-munkres.pdf>. 2004-07-15.
- [7] LABC H P. Semantic Web programme Jena 2 ontology API[EB/OL]. <http://jena.sourceforge.net/ontology/index.html>. 2004-09-30.

编辑 孙晓丹

(上接第505页)

## 3 结 论

基于模糊偏序关系的软件测试质量评价是运用模糊理论定量地说明测试方案的质量,使测试质量的评价结果具有可比性。该方法简单易行,考虑问题具有系统性,数据的获得合理科学;采用客观方法处理评价信息,减少了评价结果的误差,对测试质量的评价更有说服力。在具体应用时,可根据所具备的条件和对评价结果的要求灵活采用这种方法。

## 参 考 文 献

- [1] PAUL C J. Software testing: a craftsman's approach, [M]. 第2版. 北京: 机械工业出版社, 2003.

- [2] 王 凌, 冯 惠, 杨根兴, 等. GJB 2434A-2004, 军用软件产品评价[S]. 北京: 总装备部军标出版发行部, 2004.
- [3] 冯 惠, 王 凌, 陈森方, 等. GJB 5236-2004, 军用软件质量度量[S]. 北京: 总装备部军标出版发行部, 2004.
- [4] MORMAN E F, SHARI L P. Software metrics: a rigorous and practical approach, [M]. 第2版. 北京: 机械工业出版社, 2004.
- [5] 杜吉泽. 市场分析[M]. 北京: 经济科学出版社, 2001.
- [6] 王培庄, 韩立岩. 应用模糊数学[M]. 北京: 北京经济出版社, 1989.
- [7] 贺仲雄, 王 伟. 决策科学从最优到满意[M]. 重庆: 重庆出版社, 1988.
- [8] 黄 沛. 模糊偏序关系及其在消费者广告偏好方面的运用[J]. *中国管理科学*, 1997, 5(1): 57-64.

编辑 黄 莘