

基于 Petri 网的 OWL-S 语义匹配机制研究

裘杭萍 胡 汭 罗 晨

(解放军理工大学指挥自动化学院 南京 210007)

摘 要 传统的 UDDI 不支持语义推理以及基于服务属性的匹配,因此存在召回率低、匹配效率低等问题。基于此问题提出了一种基于 Petri 网的 OWL-S 语义匹配机制,即借助 Multi-Agent 服务发现框架,使用 PNSDL 描述语言来发布和请求服务。该匹配机制借助 Petri 网对服务的描述能力,通过使用可能性和必然性程度代表服务可以胜任需求的置信度,从而实现服务的匹配。该机制具有强大的逻辑推理能力及对多服务的建模能力,能提高匹配的真实度、匹配度等。

关键词 OWL-S 语义,语义匹配,服务发现,Petri 网,Multi-Agent

中图法分类号 TP311 **文献标识码** A

Research on Petri Net Based OWL-S Semantic Matching Mechanism

QIU Hang-ping HU Rui LUO Chen

(Institute of Command Automation, PLAUST, Nanjing 210007, China)

Abstract Traditional UDDI does not have the capability to process semantic and match to the service attribute, as a result, it has the low recall rate and matching rate. Through the service matching framework of Multi-Agent, a Petri net based OWL-S semantic matching mechanism was proposed to deal with the shortcoming, in which PNSDL description language is used to publish and request service. In virtue of the ability of Petri net for describing Web service, the matching mechanism uses the possibility and necessity to quantify the confidence levels that the service could satisfy a request, so as to finish the matching of service. The matching mechanism has the advantage of powerful ability to process, modeling for Multi-service, high reality of matching and high matching rate, etc.

Keywords OWL-S semantic, Semantic matching, Service discovery, Petri net, Multi-Agent

1 引言

近年来,云计算、一体化网络等技术迅速兴起并逐渐形成规模,社交网络、电子商务、数字城市、在线视频等新一代规模化的互联网应用发展迅猛,Web 服务的应用越来越广泛,高效的服务发现机制越来越受到关注^[1]。传统的服务发现通常是在统一描述、发现和集成 UDDI (Universal Description, Discovery and Integration)^[2] 标准基础上,通过关键字匹配和信息分类查找来实现。由于 UDDI 不支持基于服务属性的匹配和语义推理,因此存在召回率低、匹配精度不高等问题,所以增加服务属性语义匹配效率成为提高服务发现效能的有效途径。而 OWL-S (Web Ontology Language for Services)^[3] 为服务查询提供了一个语义匹配方法,其标记语言结构可以准确地刻画 Web 服务的功能和属性,能够根据服务能力和其他属性进行推理和语义匹配,进而提高服务发现的召回率和匹配精度。M. N. Huhns 提出了用语义 Web 描述语言 DAML 描述服务^[4]、用 Prolog 语言作为推理语言的服务发现系统,它根据预定义的服务属性的本体属性值来发现服务。由于在服务匹配的并发以及召回率等方面还存在缺陷,D Tras-

tour^[5] 提出以 DAML-S 语言描述服务,通过服务的属性和接口的输入输出概念匹配,得到匹配结果。这种服务匹配方式在精确度和查找效率方面还有待进一步提高。J. C. Vedral^[6] 和 J. Weber^[7] 分别提出了多种基于 P2P (Peer-to-Peer) 网络的分布式服务注册和发现模式,其关心召回率、响应时间等基本性能,解决方法是根据服务的内容,按照一定的规则或映射关系将服务存储在注册节点中,服务的内容和存储位置有一定的内在关系。该文献中提出的匹配模式相对比较完善,但是在服务匹配的描述方面还存在缺陷。本文在面向服务的 Multi-Agent 服务发现框架基础上,提出一种基于 Petri 网的 OWL-S 的语义服务发现匹配方法,其本体结构如图 1 所示。在通用服务描述模型的基础上,引入了 PN (Petri Net)^[8] 来实现语义匹配^[9,10]。通过模糊 Petri 网建立模型,运用基于 OWL-S 语义的服务匹配算法,以及数据映射、索引和近似概念搜寻算法提高服务注册和发现的匹配精度和召回率。

本文主要关注服务的召回率、查询流量负载和查找性价比 3 个度量指标。召回率 (recall) 是指查找到的目标服务数量和总的目标服务数量的比例,它反映了算法的查找能力,是服务发现算法的基本性能指标。查询流量负载 (query traffic

到稿日期:2011-12-27 返修日期:2012-03-27 本文受国家 863 基金项目——大规模复杂系统的分析能力技术(2007AA126)资助。

裘杭萍(1965—),女,博士,教授,主要研究方向为系统工程、军事通信网,E-mail:loghuhu@163.com;胡 汭(1984—),女,硕士生,主要研究方向为军事通信网。

load)是指对于一个查询请求,网络所有节点接收到的查询消息的总数量,它反映了算法消耗的网络带宽资源,可以衡量在分布式服务注册网络中服务发现算法的代价。查找性价比(search performance-price ratio)是指查找到的目标服务数量和查询消息数量的比例,它反映了平均每个消息可以找到的目标服务的数量,表现出算法的性价比。

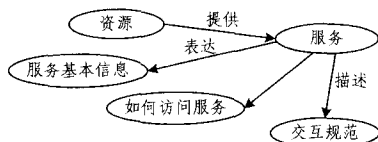


图1 OWL-S本体结构

2 服务匹配框架

虽然 UDDI 的使用有很多限制,但其作为较完备的 Web 服务注册和发现机制已被广泛应用。因此,在语义匹配模型中,服务的发现、注册和匹配沿用了 UDDI 的标准和接口。文献[8]提出了一种面向服务的 Multi-Agent 服务发现框架。本文沿用此框架,并对服务描述语言的标志和搜寻算法进行了改进。

Woodbridge 和 Jennings 给出 Agent 的定义^[11]:Agent 是一个具有自主性、社会性、反应性和能动性性质的基于硬件或软件的计算机系统。

该框架中有 3 类 Agent(如图 2 所示):服务 Agent(Service Agent),提供资源、信息、产品和领域内问题求解的能力;请求 Agent(Request Agent),服务消费者请求具体的服务;经纪人 Agent(Broker Agent),代理服务器。Agent 具有自治性、响应性、主动性、推理性以及社会性,可以在一定程度上模拟人类社会的行为,具有一定的智能,能够自主运行,有极大的灵活性和适应性,适合开放、动态的网络环境。

服务匹配的过程为:(1)服务提供者在 Service Agent 处将需要提供的资源、信息、产品等封装成服务;(2)Service Agent 对服务进行验证后,使用 Ontology 对服务进行描述,然后进行注册和发布;(3)服务请求者将请求发给 Request Agent;(4)Request Agent 在 Broker Agent 中查询 Service Agent 注册的资源,并反馈给服务请求者,而 Broker Agent 负责具体的服务发现;(5)根据反馈的结果,服务请求者使用服务提供者发布的服务。

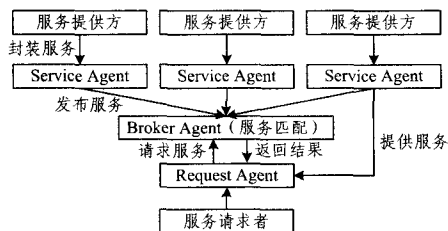


图2 基于 Petri 网的 OWL-S 语义匹配框架

相对于传统的服务注册-发布-请求机制,Multi-Agent 机制将服务的发布、请求和查询分别从服务提供方、服务请求者和注册中心转移到了 Service Agent、Request Agent 和 Broker Agent,虽然增加了可以忽略不计的系统开销,但简化了终端的设计,并且在 Broker Agent 完成服务发现后,可以有效地提高召回率、查找性价比和降低查询流量负载。

在框架中,核心机制是服务匹配,而服务匹配的实现在

Broker Agent 中完成的。Broker Agent 接收到服务发布方提供的服务 advertisements 和请求者的 inquiries 之后,PNSDL 通过扩充 Tmodel 信息,建立一个 UDDI 服务能力描述,生成该服务的一个索引 ID,并将其传给外部 UDDI 注册中心。同时,在 Broker 内部,利用本体,通过逻辑推理来进行服务的语义匹配,从而实现有效的服务发现。

3 基于 Petri 网的服务描述语言

3.1 Petri 网介绍

Petri 网是一种适合于描述异步并发现象的系统模型,起源于 Carl Adam Petri 1962 年的博士论文。Petri 网既有严格的数学定义,又有直观的图形表示。其基本构成为库所(place)、变迁(transition)和弧(arc)。库所描述可能的系统局部状态,变迁描述触发系统局部状态改变的事件,弧用来连接库所和变迁。库所到变迁的弧表示事件能够发生的局部状态,变迁到库所的弧表示由事件所引发的局部状态的转换。

Petri 网的形式化定义如下:

Petri 网的定义为:一个三元组 $N=(S, T, F)$ 是一个 Petri 网,当且仅当:

- (1) $S \cup T \neq \emptyset$ (非空性);
- (2) $S \cap T = \emptyset$ (二元性);
- (3) $F \subseteq (S \times T) \cup (T \times S)$ (流关系仅在于 S 与 T 的元素之间);

(4) $dom(F) \cup cod(F) = S \cup T$ (没有孤立元素)。其中, F 的元素叫弧;

$$dom(F) = \{x \mid \exists y: (x, y) \in F\};$$

$cod(F) = \{x \mid \exists y: (y, x) \in F\}$ 。集合 $X = S \cup T$ 是网元素的集合。

本文使用 Petri 网作为服务描述语言的基础,并称其为 PN-SDL (Petri Net Service Description Language)。采用 OWL-S 表示共享领域知识。PNSDL 是指导 Agent 在开放的网络环境下发布、请求和匹配服务的规范:

1) 推理变迁:在 Petri 网中,变迁表示前提条件成立所执行的事件或动作,即执行变迁后产生的后果。

2) 输入输出变量:推理变迁的输入弧上的变量是实际需要匹配的输入变量;而输出弧上的变量是实际需要匹配的输输出变量。

3) 输入库所表示服务执行前输入变量成立;输出库所表示服务执行后输出变量成立。

3.2 PNSDL 本体描述

对 Petri 网进行拓展,定义能够进行服务描述的 Petri 网为六元组: $PN=(P, T_p, F, W, M_0, CF)$, 其中 $P=\{p_1, p_2, \dots, p_m\}$ 是库所的有限集; $T=\{t_1, t_2, \dots, t_m\}$ 为有限变迁集; $F \subseteq (P \times T) \cup (T \times P)$ 表示一个弧集(即流关系); $W: F \rightarrow \{1, 2, \dots\}$ 表示一个权函数; $M_0: P \rightarrow \{0, 1, 2, \dots\}$ 是一个初始标识; θ 为真实度向量, $\theta=(\theta_1, \theta_2, \dots, \theta_n)$, 其中 $\theta_i \in [0, 1]$ 表示命题 p_i 的真实度, $i=1, 2, \dots, n$; θ' 为初始真实度向量; $CF=\{u_1, u_2, u_3, \dots, u_m\}$, u_j 是规则 r_j 成立的置信度, $j=1, 2, 3, \dots, m$; u_j 的值介于 0, 1 之间,数值越大置信度越高。

可借助 OWL-S 来实现 PNSDL 的功能。设计的服务描述语言的信息主要是服务本体。在 Broker Agent 中,服务本体的基本描述如图 3 所示。

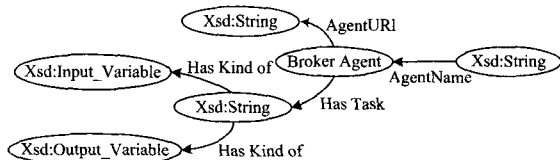


图3 基于 PNSDL 的服务本体

服务本体的主要作用是描述服务是做什么的和如何访问服务。Agent 基础主要有提供服务 Agent 的 URL、名称、任务和标识(token)。服务 Agent 通过这些标识来体现发布服务和执行任务。

4 基于 OWL-S 的服务匹配机制

4.1 相似度和蕴含度

为了实现基于 OWL-S 的服务匹配,根据对 PN 六元组的定义,服务匹配的设计需要对类之间的置信度进行量化,也就是量化其可能性和必然性。因此首先需要对相似度、一致性和蕴含性进行定义。

而本体(ontology)定义了用户在共享信息时所使用的一种通用词汇,用以定义领域内的概念、概念与概念间的相关性。在 Multi-Agent 的框架下,通过分层(hierarchy)的方法,建立类之间的关系,以发现语义上的关联性。

定义 1(邻接的超类和子类的相似度) 对两个邻接的超类和子类(c_1, c_2) 都赋予一个惟一的相似度 $S(c_1, c_2)$, 范围从 0 到 1。

定义 2(两类的最小公共超类) 令类 c_1 和 c_2 的最小公共超类记为 $CS(c_1, c_2)$, 有:

$$CS(c_1, c_2) = \{g | su(g, c_1) \wedge su(g, c_2) \wedge (\forall c \neq g)(su(c, c_1) \wedge su(c, c_2) \rightarrow su(c, g))\}$$

其中, $su(x, y)$ 是谓词, 表示 x 是 y 的超类。 $CS(c_1, c_2)$ 指出了包含 c_1 和 c_2 的最特殊的类。如果 c_1 和 c_2 两个类位于两个不同的类分层中, 则 $CS(c_1, c_2)$ 不存在。

定义 3(任意两个类的相似度) 令类 c_1 和 c_2 的相似度记为 $S(c_1, c_2)$, 有:

$$(1) S(c_1, c_2) = S(c_1, CS(c_1, c_2)) \times S(c_2, CS(c_1, c_2));$$

$$(2) S(c_1, c_2) = 1, \text{ 如果 } c_1 = c_2;$$

$$(3) S(c_1, c_2) = 0, \text{ 如果 } CS(c_1, c_2) \text{ 不存在。}$$

定义 4(两个类的一致性) 令类 c_1 和 c_2 的一致性记为 $C(c_1 \rightarrow c_2)$, 有:

$$(1) C(c_1 \rightarrow c_2) = 1, \text{ 如果 } CS(c_1, c_2) \text{ 存在};$$

$$(2) C(c_1 \rightarrow c_2) = 0, \text{ 如果 } CS(c_1, c_2) \text{ 不存在。}$$

定义 5(两个类的蕴含性) 记类 c_1 蕴含 c_2 的程度为 $I(c_1 \rightarrow c_2)$, 有:

$$(1) I(c_1 \rightarrow c_2) = 1, \text{ 如果 } CS(c_1, c_2) = c_2;$$

$$(2) I(c_1 \rightarrow c_2) = S(c_1, c_2), \text{ 如果 } c_1 \text{ 和 } c_2 \text{ 在同一个类分层中且 } CS(c_1, c_2) \neq c_2;$$

$$(3) I(c_1 \rightarrow c_2) = 0, \text{ 如果 } c_1 \text{ 和 } c_2 \text{ 不在同一个类分层中。}$$

4.2 匹配算法

假设有一个服务与一个请求, 请求有 m 个输入库所来描述在没有得到服务之前的情况, 用以下的算法选择一个有 n 个输入库所的服务去匹配需求($n \leq m$)。

匹配算法:

(1) 如图 4 所示, 在本体图中, 基于语法与语义信息, 查

找服务的输出库所信息 $OP_1(q_1)$ 和请求的输出库所信息 $OP_2(q_2)$ 。

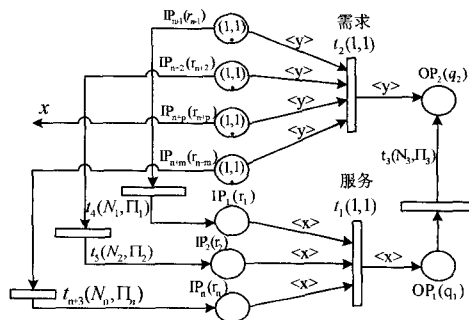


图4 服务匹配需求示意图

(2) 根据类的蕴含性, 如果类 c_{q_1} 和 c_{q_2} 表现为位于两个不同的类分层中, 它们就属于不同的本体, 则选择另外一个服务并转步骤(1)。

(3) 引入推理变迁 t_3 的输入库所是 $OP_1(q_1)$, 输出库所是 $OP_2(q_2)$, 计算 t_3 的置信度, 亦即对服务 Agent 满足所有前提条件的置信度; 计算服务的输出库所能推出需求的输出库所的可能性 $\Pi = C(c_{q_2} \rightarrow c_{q_1})$ 和必然性 $N = I(c_{q_2} \rightarrow c_{q_1})$ 。

(4) 计算满足服务所需前提条件的置信度: 在类分层中找到需求的输入库所 $IP_1(r_1)$ 和服务的输入库所 $IP_k(r_k)$ 中提到的类 c_{r_1} 和 c_{r_k} 的位置; 如果 c_{r_1} 和 c_{r_k} 有相同的类分层, 则计算需求的输入库所能推出服务的输入库所的可能性 $\Pi(c_{r_1} \rightarrow c_{r_k})$ 和必然性 $N(c_{r_1} \rightarrow c_{r_k})$ 。

(5) 如果任何一对 (c_{r_1}, c_{r_k}) 都不在相同的本体中, 即, 当前的服务不满足需求, 则选择另外的服务并转步骤(1)。

(6) 根据步骤(4)中推出的可能性和必然性程度, 在需求的所有输入库所中插入确定的标识(token), 即 $(N, \Pi) = (1, 1)$ 作为当前的前提条件, 然后通过变迁 $t_n \sim t_{n+3}$ (其输入库所是需求的输入库所, 输出库所是服务的输入库所) 来执行可能性推理。

(7) 通过确定的服务变迁 t_1 来执行可能性推理, 然后通过变迁 t_3 来执行可能性推理; 最后导出的可能性和必然性程度就是对该服务能实现需求的总体置信度。

(8) 引发变迁 t_1, t_3 来执行模糊推理, 从而求出命题的真实度, 即服务可以实现匹配的程度真实度说明服务满足请求的程度。

5 实验分析

为了检验本文所提出的 PNSDL 描述语言及其匹配机制在实践应用中的可行性和有效性, 实验在一台 Intel E7450 (2400GHz) 处理器、内存 4GB 的 DELL R900 服务器上进行, 操作系统为 Microsoft Windows XP, UDDI 环境为 jUDDIv 2.0, OWL 推理工具为 Pellet。根据评估服务匹配系统的机制和匹配算法, 对服务搜索结果的准确度、召回率进行实现和记录。实验使用 Protégé 软件设计了若干 OWL 本体, 并从互联网下载和选择了一批本体。利用这些本体, 仿真了 1000 个服务的 OWL-S 描述和对应的 UDDI 数据格式的描述。实验系统把请求的语义属性条件集和关键字条件集分别封装为 OWL-S 服务发现请求和 UDDI 服务发现请求。设定不同的 PNSDL 的真实度, 观察在不同 PNSDL 的真实度下, 两种服

(下转第 213 页)

- [12] Sarah Y, Salem F, Duan Yang, et al. The evolution of transmembrane helix kinks and the structural diversity of g protein-coupled receptors[Z]. California Institute of Technology, Pasadena, CA, 2003; 959-963
- [13] Best D J, Fisher N I. Efficient simulation of the von mises distribution[J]. Journal of the Royal Statistical Society, Series C (Applied Statistics), 1979, 28(2): 152-157
- [14] Meruelo A D, James I S, Bowie U. TMKink: A method to predict transmembrane helix kinks[J]. Protein Science, 2011, 20: 1256-1264
- [15] Apweiler R, Bairoch A, Wu C. Uniprot: the universal protein knowledgebase[J]. Nucleic acids Research, 2004, 32: 115-119
- [16] Zhao Ying, Karypis G. Evaluation of hierarchical clustering algorithms for document datasets[C]//Proc. of Int'l. Conf on Information and Knowledge Management, 2002; 515-524

(上接第 176 页)

务描述方式对应的召回率和准确度的函数关系。在评价一个匹配算法的效率时,片面地考虑召回率或准确率都不够全面。为了全面地考察实验结果,在准确率为 Y 轴、召回率为 X 轴的图上画出了不同真实度下的准确率与召回率的函数关系图形。实验统计结果见图 5,两条曲线分别对应了两种匹配算法的召回率-准确率曲线。不难看出,在大多数情况下星点实线所标示的性能曲线性能高于虚线所标示的性能曲线。不难发现一个匹配算法的性能越好,其对应的召回率-准确率曲线越向上凸,具体来讲就是曲线与 X 轴、Y 轴所围成的图形的面积越大,匹配算法的性能越好。最理想的情况是,面积为 1。在不同真实度条件下,召回率和准确率统计实验结果显示了基于 Petri 网的 OWL-S 语义匹配相对于传统 UDDI 匹配机制具有明显优势。

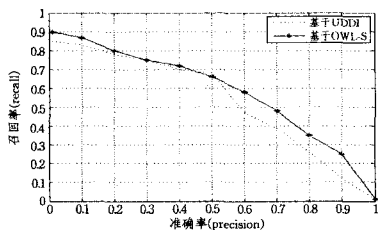


图 5 实验结果

结束语 本文提出了一种基于 Petri 网的 OWL-S 语义匹配机制,即借助 Multi-Agent 服务发现框架,使用 PNSDL 描述语言来发布和请求服务。提出了一个基于 OWL-S 语义的服务匹配机制,给出了相应的匹配算法。该匹配机制以用可能性和必然性程度代表服务能胜任需求的置信度,实现服务的匹配以提高服务发现的召回率和准确度的效率。该机制具有强大的逻辑推理能力,并能对多服务建模,提高匹配的真实度、准确度和召回率。

下一步将在 Multi-Agent 服务发现框架及大型的分布式环境中对服务组合问题进行研究。

参考文献

- [1] 岳昆,王晓玲,周傲英. Web 服务核心支撑技术:研究综述[J]. 软件学报,2004,15(3): 428-442
- [2] Clement L, Hatley A, von Riegen, et al. UDDI Version 3. 0. 2

- [17] 黄旭,吕强,钱培德. 一种用于蛋白质结构聚类的聚类中心选择算法[J]. 自动化学报,2011,37(6): 682-692
- [18] von Luxburg U. A tutorial on spectral clustering[J]. Statistics and Computing, 2007, 17(4): 395-416
- [19] Mardia K V, Taylor C C, Subramaniam G K. Protein bioinformatics and mixtures of bivariate von mises distributions for angular data[J]. Journal of the Royal Statistical Society, Series C (Applied Statistics), 2007, 63(2): 505-512
- [20] Fisher N I. Statistical Analysis of Circular Data[M]. Cambridge University Press, 1996
- [21] wikipedia [EB/OL]. [http://en.wikipedia.org/wiki/Talk:Von Mises distribution](http://en.wikipedia.org/wiki/Talk:Von_Mises_distribution). URL, Oct. 2010
- [22] Rasmussen M, Karypis G. gcluto: An interactive clustering, visualization, and analysis system[R]. UMN-CS TR-04-021. 2004
- [23] Deupi X, Olivella M, Govaerts C, et al. Ser and thr residues modulate the conformation of pro-kinked transmembrane α -helices [J]. Biophysical Journal, 2004, 86: 105-115
- [EB/OL]. [http://uddi.org/pubs/uddi2v3. 0. 2-20041019. htm](http://uddi.org/pubs/uddi2v3.0.2-20041019.htm), 2004
- [3] Martin D, Burstein M, Hobbs J, et al. OWL2S: Semantic Markup for Web Services[EB/OL]. [http://www.w3.org/ Submission/ 2004/ SUBM2OWL2S220041122/](http://www.w3.org/Submission/2004/SUBM2OWL2S220041122/), 2004
- [4] Huhns M N, Singh M P. Ontologies for agents [J]. IEEE Internet Computing, 1997, 1(6): 81-83
- [5] Trastour D, Bartolini C, Castillo J G. A service Web approach to service description for matchmaking of services [R]. HP Labs, Tech Rep; HPL2001-183. 2001
- [6] Verma K, Sivashanmugam K, Sheth A, et al. METEOR-S WSDI: A scalable P2P infrastructure of registries for semantic publication and discovery of Web services[J]. Journal of Information Technology and Management, 2005, 6(1): 17-39
- [7] Chen C W, Gan P S, Yang C H. A service discovery mechanism with load balance issue in decentralized peer-to-peer network[C]// Barolli L, ed. Proc. of the 11th Int'l Conf. on Parallel and Distributed Systems (ICPADS 2005). Washington: IEEE Computer Society Press, 2005: 592-598
- [8] 马千里, 廖明宏, 高振国, 等. 普适计算环境下的服务发现协议 [J]. 计算机工程, 2009, 35(17): 247-248
- [9] Petri C A. Kommunikation mit Automaten[D]. Bonn, Germany: Institute for Instrumentelle Mathematik, Schriften des IIM, 1962
- [10] Vedral J C, Lama M, Bugarin A. A High-level Petri Net Ontology Compatible with PNML[C]// ACSD, Petri Net Markup Language Forum 2006. Turku, 2006
- [11] Weber J, Kindler E. The Petri Net Markup Language[J]. Lecture Notes in Computer Science, 2003, 2472: 124-144
- [12] 张广胜, 蒋昌俊, 丁志军. 基于模糊 Petri 网的服务发现框架研究 [J]. 计算机研究与发展, 2006, 43(11): 1886-1894
- [13] Wooldridge M. Intelligent Agent: Theory and Practice [J]. Knowledge Engineering Review, 1995, 10(2): 115-152
- [14] 郑啸, 罗军舟, 宋爱波. 基于 Agent 和蚁群算法的分布式服务发现 [J]. 软件学报, 2010, 21(8): 1795-1809
- [15] 翟正利, 杨扬. 基于模糊 Petri 网和本体的网格服务发现 [J]. 北京科技大学学报, 2006, 28(12): 1196-1200
- [16] 邱田, 胡晓惠, 李鹏飞, 等. 基于 OWL-S 的服务发现语义匹配机制 [J]. 电子学报, 2010, 38(1): 42-47