

# 基于信誉的语义 Web 服务发现<sup>\*</sup>)

王 慧<sup>1</sup> 王金华<sup>2</sup> 赵焯辉<sup>1</sup> 高 远<sup>1</sup> 王翠荣<sup>1</sup>

(东北大学信息科学与工程学院 沈阳 110004)<sup>1</sup> (中国网通集团秦皇岛分公司)<sup>2</sup>

**摘 要** 服务发现在语义 Web 领域是一个挑战性问题。一个关键问题是发现满足用户功能需求的最佳服务。同样重要的是,电子商务应用还要求发现满足 QoS 需求的最佳服务。由此提出一个基于信誉管理的分布式服务选择框架。它主要有以下特点:用一个结构化的覆盖网络作为分布式服务存储系统来提高系统的有效性和可扩展性;提出一个 QoS 感知的语义 Web 服务发现方法,并给出基于用户反馈的分布式信誉度评估算法。最后用一个例子示范算法如何解决实际问题。

**关键词** 语义 Web 服务, SOA, QoS, 信誉

## Reputation-based Semantic Web Service Discovery

WANG Hui WANG Jin-Hua ZHAO Yu-Hui GAO Yuan WANG Cui-Rong  
(School of Information Science & Engineering, Northeastern University, Shenyang 110004)

**Abstract** Service discovery is a challenging problem in the Semantic Web service area. One key issue is to discover the most relevant services meeting the functional requirements of users. Equally important, e-Business applications also would like to discover services which best meet their requirements in terms of QoS. In this paper, we present a framework to facilitate Reputation-based service selection, which has three key features. First, we use a structured overlay network as the decentralized service repository system. Second, we propose an adaptive QoS-aware service discovery approach. Then, we present a service oriented distributed reputation assessment algorithm as a QoS support module in our framework. We test our approach on a simple, yet realistic example, and the preliminary results demonstrate that our implementation provides a practical solution.

**Keywords** Semantic Web service, SOA, QoS, Reputation

面向服务的体系结构(SOA)<sup>[1]</sup>,是分布式计算领域中采用的一种主要体系结构。在这个架构中,服务提供者提供一些(潜在竞争的)服务,服务请求者根据本身的需求动态的选择最佳的服务。SOA 使服务能够在运行时被动态的选择和整合,因此使系统具有灵活性并适应现代商务必须具备的自治特性。Web 服务组合强调两个问题。首先,可扩展的 Web 服务的发现机制是基础,它为用户提供可选择的 Web 服务,满足它们的功能需求和 QoS 需求。其次,信誉管理机制是补充,使之可基于服务提供者是否提供了可靠服务和用户是否提供了真实的反馈报告来预期未来服务的质量。

本文提出一个分布式服务发现方法并用结构化覆盖网络作为分布式的服务注册系统以提高系统的可扩展性和服务的效率。我们假定 Web 服务用包含 QoS 属性的语义描述,并存储在分布式的服务代理节点(SvBs)上,用户可根据服务提供者实际提供的服务质量反馈评估报告。基于以上这些假设,我们提出基于信誉管理的 Web 服务选择算法,用信任和信誉评估技术来预期未来的服务质量。它由两部分组成,一是用户报告预处理阶段,二是信誉度评估阶段。

## 1 相关工作

服务发现在语义 Web 领域是一个挑战性问题。一个关键问题是发现满足用户功能需求的最佳服务。同样重要的

是,电子商务应用还要求发现满足 QoS 需求(例如,性能、吞吐量、可靠性、有效性等)的最佳服务。因此基于 QoS 的 Web 服务选择机制将在 SOA 中起重要作用,特别是当语义匹配过程返回带有可比较其功能的多个服务时。尽管传统的 UDDI 标准没有提供对 Web 服务的 QoS 描述,许多建议已经被提出扩展最初的模型,并描述 Web 服务质量能力<sup>[2]</sup>。文[3]提出对 UDDI 标准的一个扩展的补充,存储由服务提供者或消费者提交的 QoS 数据。文[4]提出根据用户评定的服务等级、依据服务质量以及服务质量的真实性三个要素评估服务的信誉度。然而,这些解决方法并没有考虑由用户生成的可信 QoS 报告,而它是确保基于 QoS 的精确选择 Web 服务的重要保证。

信誉管理系统已经被广泛应用于电子商务系统,当两个未知的对等节点互相交易时,它作为一个有效的机制来处理可能出现的风险。在 Web 服务合成中,作为对服务过程管理和发现机制的补充,已有各种信誉模型被提出。文[5]给出一个基于 QoS 的选择模型,它考虑来自用户的反馈和其它与交易相关的因素,其中服务等级的计算是通过监控服务和用户取得 QoS 的执行情况得出的。文[5]和[7]提出一个基于 P2P 覆盖网的 Web 服务选择和信誉管理方法。文[8]建议增大对服务客户的 QoS 监控、分析和选择能力,但这是不切合实际的,因为每个服务消费者可能必须负担沉重的服务发现

<sup>\*</sup> 基金项目:东北大学“985 工程”基金(985-2DC-C09)。王 慧 博士研究生;赵焯辉 博士研究生;高 远 教授,博士生导师;王翠荣 教授。

和信誉管理的角色。我们在当前存在工作的基础上, 给出一个基于 QoS 的服务发现模型, 并提出一个根据可信用户的反馈信誉度评估算法。与此领域的现存工作相比, 我们的工作 在提高系统的扩展性和灵活性方面更有优势。

## 2 系统结构

### 2.1 服务发现模型

本文中的服务发现模型是一个基于 P2P 覆盖网络拓 扑结构的分布式服务合成环境。图 1 展示了分布式服务发现框 架的概念模型。

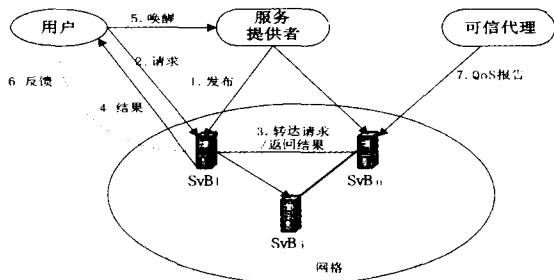


图 1 服务发现模型

(1) 服务提供者在覆盖网络中的服务代理节点 SvBs 上发 布含有 QoS 信息的 Web 服务。(2) 用户可将任何 SvBs 作为 访问入口请求有确定功能并包含 QoS 属性的服务。(3) SvBs 负责路由这个服务请求到能提供这个服务的 SvBs 上。(4) 用 户唤醒被发现的服务。(5) 结果返回给用户。(6) 用户反馈给 提供服务的 SvBs 关于它所获得服务的 QoS 性能。此 SvBs 评估用户的 QoS 报告, 检查是否有恶意报告和用户的共谋欺 诈行为, 以获得对 QoS 服务的正确评价。(7) 信任代理提供 服务质量监控报告。我们允许信任代理模式在系统中对特定 的服务提供 QoS 监控。这些权威机构的信任代理总是提供 可靠的 QoS 报告并作为可信赖的信息源来评估其它用户的 行为。在本模型中, 我们并不配置这些代理收集所有被登记 的有效服务的执行数据, 而是仅仅监控被选择服务的服 务质量, 因为这些特殊代理的启动和维护通常需要很高的费用。

### 2.2 服务代理内部结构

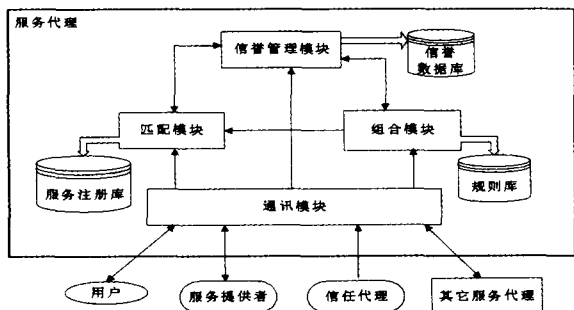


图 2 服务代理的内部结构

图 2 展示了一个服务代理的内部结构, 它包括通讯管理 模块、匹配模块、组合模块和信誉管理模块以及存储相关信 息的服务注册库、信誉数据库和服务组合中使用的规则库。

通讯管理模块提供一个信息总线连接系统内其它内部 模块, 并与系统外部实体交互, 例如用户、信任代理、服务提供 者, 用以获得服务广告、QoS 数据以及用户反馈, 并将这些信 息提供给内部服务模块。另外它还是 SvBs 与其它对等代理 节点的接口。它使代理节点之间能互相传播、交换注册的服

务信息和 QoS 数据。同时它接受用户提交的服务请求, 请求 匹配模块找到一个包含 QoS 需求的语义匹配的原子或组合 服务并返回执行结果。

匹配模块将用户请求的服务功能和相应的 QoS 需求转 化为一个语义 Web 服务请求, 并与服务注册库中的服务描述 匹配。同时匹配模块也请求信誉管理模块选择满足 QoS 需 求的最佳服务。匹配模块确保所发布服务的 QoS 属性满足 服务请求的要求。我们扩展由 Paolucci 等<sup>[9]</sup>提出的 DAML-S 匹配算法, 增加了 QoS 匹配功能。算法用于对一个请求服务 与一系列发布的服务进行匹配。请求服务有一系列规定的输 入  $in_{Req}$  和预期的输出  $out_{Req}$ , 反之, 每个发布的服务有一个 预期的输入  $in_{Adv}$  和规定的输出  $out_{Adv}$ 。这个匹配算法在两个 本体概念(一个请求的概念  $C_{Req}$  和一个发布概念  $C_{Adv}$ )之间定义 了四层的匹配关系。

- 精确匹配: 如果  $C_{Req} = C_{Adv}$  或者  $C_{Req}$  是  $C_{Adv}$  的一个直接子集。
- 插入式匹配: 如果  $C_{Adv}$  包含  $C_{Req}$ , 在这种情况下,  $C_{Adv}$  可以被  $C_{Req}$  使用。
- 包容匹配: 如果  $C_{Req}$  包含  $C_{Adv}$ , 在这种情况下, 服务不 能完全满足需求, 因此可能需要其它服务来满足其余的服务 需求。
- 失败匹配: 当在发布概念和请求概念之间没有相容关 系时, 匹配失败。

我们在两个方面修改了这个算法。首先, 我们在语义匹 配之后, 进行数据类型的兼容性检查。然后, 服务选择必须能 够对所请求的 QoS 进行匹配。

组合模块试图建立一个互相连接在一起实现用户目标 的组合服务链。在匹配模块不能找到单一的 Web 服务来满足 请求的功能时, 通讯管理模块请求组合模块将一些服务组合 在一起来满足请求的服务功能并同时满足 QoS 需求。

组合模块利用一个动态的适应算法, 通过至少两个不 同的信息资源, 尝试重新找到匹配的服务。通过查询一个提供 组合模板的规则库(Rule DB)查找与服务请求相匹配的规则。 根据找到的规则, 组合模块试图对这个规则的相应模板中每 个元素的输入输出与服务注册库(SR)中的 Web 服务进行语 义匹配。如果匹配失败, 组合模块将试图寻找另一个规则, 它 可以进一步分解这个模板。如此这般, 循环往复, 直至找到一个 原子规则。最终, 服务被连接在一起形成符合特定规则的一 个工作流图。

信誉管理模块计算每个服务的信誉度, 并根据用户的需 求返回有最高级别信誉度的服务。提供者能查询所提供服 务的信誉度并决定是否改进服务的执行能力。其细节将在第 3 部分加以描述。

## 3 信誉评估算法

### 3.1 预处理阶段

报告预处理方法是依靠一些来自特殊监控代理的信任 报告, 应用“信任-不信任”传播方法, 评估所收集用户报告的 可信性。如果在一个特定的时间段  $\Delta$  内, 两个 QoS 报告联系 到同一个服务, 我们认为它们是可比较的, 否则是不可比较 的。通常, 只要相应的服务提供者在此期间不改变这个服 务的 QoS 允诺值, 我们可设  $\Delta$  的时间长度尽可能大。如果它 们之间一致性值  $c_i$  距离的平方(我们称其为它们之间的相似 度)比一个特定的门限值小的话, 两个可比较的 QoS 报告被

认为是相似的。反之,如果它们之间的距离比另一个门限值大的话,则认为是不相似的。

我们利用报告预处理算法对报告进行预处理,根据它们与其它诚实或欺诈报告之间的相似度评估用户是否诚实。 $n_{ch}$ 、 $n_{h1}$ 和 $n_{h2}$ ( $n_{h1} < n_{h2}$ )是门限值,它们是设计参数,在运行此算法几次之后,由所收集报告的特性确定。

报告预处理算法如下:

步骤1 标注所有可信用用户代理和它们的报告

- (1)所有可信用用户代理被标注为诚实用户。
- (2)所有可信用用户代理的报告被标注为诚实的报告。

步骤2 标注所有欺诈用户和欺诈报告

- (1)每个欺诈用户的所有未标注的报告被标注为欺诈。
- (2)对每个未标注的用户,如果其报告与欺诈报告的相似度小于 $n_{ch}$ ,则标注为欺诈。

(3)对每个未标注的报告,如果这个报告与诚实报告不相似,则标注它为欺诈报告。如果这个报告与欺诈报告相似,也标注它为欺诈。

(4)重复此步骤,直到没有新的欺诈用户被发现。

步骤3 标注所有诚实用户和诚实报告

- (1)每个诚实用户所有未标注的报告被标注为诚实。
- (2)对每个未标注的用户,如果其报告与诚实报告的相似度小于 $n_{h1}$ ,则标注为诚实。

(3)对每个未标注的报告和每个已标注为欺诈的报告,如果这个报告与一个诚实报告相似,则标注为诚实。

(4)被标注为欺诈的用户并且它的报告与诚实用户的报告的相似度小于 $n_{h2}$ ,可被重新标注为诚实。

(5)重复此步骤直至无新的诚实用户被发现。

### 3.2 信誉度计算阶段

在完成预处理阶段之后,我们可以识别出大部分的欺诈用户和诚实用户。然后根据诚实用户的反馈报告计算信誉值。

为了预计一个Web服务S的质量,我们收集在一个时间段中所有服务执行情况的信誉反馈,并用一个实际价值时间系列预报技术根据过去的的数据预计未来服务质量的一致性。

**定义1** 服务S的质量一致性值 $c_i$ 在时间 $t$ 提供的一个质量属性 $q_i$ 被定义为:

$$c_i = \frac{p_i - d_i}{p_i}$$

其中 $d_i$ 是S在时间 $t$ 实际交付给一个特定用户的标准质量 $q_i$ 值,并且 $p_i$ 是相应的服务提供者许诺提供的标准QoS值。

**定义2** 信誉管理器接受关于一个服务的信誉询问请求,我们定义,在一个用户请求中的信誉需求为一个三元组 $\{q_i, \eta_i, v_i\}$ ,其中 $q_i$ 代表QoS属性, $\eta_i$ 是属性 $q_i$ 对用户的重要性的级别(权重)。 $v_i$ 是 $q_i$ 的值。图4显示了一个带有QoS的服务请求实例。对一个服务的特定QoS属性的信誉值的计算被定义为 $q_i$ 信誉等级的平均值:

$$R_i = \frac{\sum_{j=1}^n r_{ij}}{n}$$

其中, $R_i$ 是服务S关于QoS属性 $q_i$ 的信誉值, $r_i$ 是S关于 $q_i$ 的信誉等级,而 $n$ 是评估者的数目。

**定义3** 带有多个QoS属性的服务信誉值为各个质量属性的信誉值的权重的总和。

$$R(s) = \sum_{i=1}^n \eta_i \times R_i$$

其中 $R(s)$ 是服务S所有质量属性( $q_i, q_{i+1}, \dots, q_{i+n}$ )的信誉

值。

我们定义 $r_i = \omega_i \cdot D_i$ 。其中 $D_i = (d_i - v_i) / v_i$ 评估服务S预期能够提供给用户的质量属性 $q_i$ 的值 $d_i$ 与用户请求的相应值 $v_i$ 之间的差别。结合定义1,我们可计算 $D_i = \frac{(1 + p_i)c_i - v_i}{v_i}$ ,其中 $c_i$ 是被预期的质量属性 $q_i$ 的QoS一致性值。 $p_i$ 是服务提供者许诺的在当前时间服务S提供的相应QoS值。 $\omega_i$ 是权重比例,根据 $q_i$ 与QoS本体概念的语义匹配度 $m_i$ 决定其权重。匹配度在[9]中被阐述。在本文中,我们简单的使用下面的定义来评估 $\omega_i$ :

If  $m_i$  = 精确匹配  $\omega_i = 1.0$

If  $m_i$  = 插入式匹配  $\omega_i = 0.5$

If  $m_i \in \{ \text{包容匹配}, \text{失败匹配} \}$   $\omega_i = 0.0$

为了加速所有请求QoS参数的满足需求的服务选择,我们用信誉数据库中的一个信誉等级表来存储所有频繁存取的QoS信息。对表中的每个 $q_i$ ,我们有一个记录 $\{S_i, R_i\}$ 的列表L,其中 $R_i$ 在前面已经被计算,而 $S_i$ 标识当前支持 $q_i$ 的服务。给定服务列表L,信誉管理器基于用户的请求计算每个服务的信誉值并返回有最高信誉的服务。

## 4 实例介绍

图3显示了一个典型的轿车本体描述片断。这里展示的例子是Paolucci等[9]中所描述例子的一个变种,一个请求代理要查找一个提供QoS保证的丰田轿车的销售服务。请求代理发送一个包含特定QoS参数的DAML-S格式的服务请求,在这种情况下,请求必须识别本体中一个特定的类别和这个类中特定的用户喜好的QoS属性(在例子中这些包括可靠性和有效性)。

假定服务注册库中有三个满足本体DAML规范的有效服务:S1销售丰田轿车,S2也销售丰田轿车,S3销售所有类型的轿车。输入和输出被概括在表1中,假定输入对所有服务是相同的:Concepts, daml# Price. 输出尽管不同,但假定来自相同的命名空间:Vehicle, daml.

表1 服务广告输出类型

Service	S1	S2	S3
Output	# ToyotaSadan	# ToyotaSadan	# Sadan

假定服务S1,S2,S3中的信誉参数 $R_i$ 被定义在表2中,并且它相对可靠性、有效性的权重分别是0.9和0.8。根据定义3,S1,S2,S3的信誉值可分别被计算为:S1=0.9×0.7+0.8×0.5=1.03,S2=0.9×0.3+0.8×0.2=0.43和S3=0.9×0.5+0.8×0.6=0.93

表2 信誉等级表

Service	S1	S2	S3
Availability	0.7	0.3	0.5
Reliability	0.5	0.2	0.6

在这个场景中,因为S1,S2,S3是语义相似的,通讯管理器需要在它们中间基于信誉值进行选择。

第一选择:S1-因为有精确的功能匹配和最高的信誉匹配。

第二选择:S3-因为有插入式的功能匹配和比较高的信誉匹配。

第三选择: S2-因为有精确的功能匹配和比较低的信誉匹配。

因此, 通讯管理器将返回 S1 作为最佳的匹配。

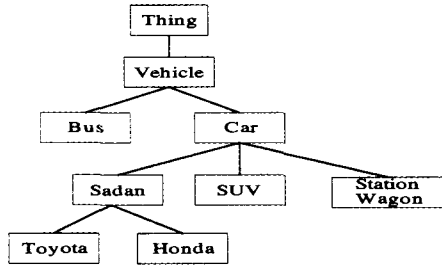


图 3 一个轿车本体

```

<profile:Profile rdf:ID="RequestToyotaSellService">
  <input>
    <profile:ParameterDescription rdf:ID="Price_Input">
      <profile:parameterName>Price</profile:parameterName>
      <profile:restrictedTo rdf:resource="Concepts.daml#Price">
    </profile:ParameterDescription>
    </input>
  <output>
    <profile:ParameterDescription rdf:ID="Car_Output">
      <profile:parameterName>ToyotaSedan</profile:parameterName>
      <profile:restrictedTo rdf:resource="Vehicle.daml#ToyotaSaloon">
    </profile:ParameterDescription>
    </output>
  <rns:ReputationAttributes>
    <rns:attribute>reliability</rns:attribute>
    <rns:attribute>availability</rns:attribute>
  </rns:ReputationAttributes>
</profile:Profile>
    
```

图 4 包含信誉参数的服务请求例子

**结束语** 本文用一个结构化的覆盖网络作为分布式的服务注册系统来提高其可扩展性和有效性。我们提出一个服务发现模型, 利用语义 Web 技术促进基于信誉的服务发现和匹配。框架包括通讯管理模块、匹配模块、组合模块、信誉模块。

文中详细描述了这些组件并提出了一个基于用户反馈的信誉评估算法。最后通过一个例子展示了此算法的使用。比较此领域已做的工作, 该模型的优点是服务信誉度加入了语义信息, 并综合了第三方权威机构的评价, 使得信誉度值更有效和准确。

本文的工作是我们正在进行研究工作的一部分, 目的在于为服务组合机制提供一个开放的、可重用的基础结构。今后, 我们计划进一步改进我们的服务发现模型, 特别是在复杂环境下的信誉管理方案, 并在真实的 Web 服务整合环境下配置这个模型, 考察它的实用性。

参考文献

- 1 Papazoglou M P, Georgakopoulos D. Service-oriented computing. Communications of the ACM, 2003, 46
- 2 Bilgin A S, Singh M P. A DAML-Based Repository for QoS-Aware Semantic Web Service Selection. In: Proceedings of ICWS'04
- 3 Kalepu S, Krishnaswamy S, Loke S W. Reputation = f(User Ranking, Compliance, Verity). In: Proceedings of ICWS'04
- 4 Singh M P, Huhns M N. Service-Oriented Computing. Wiley, 2005
- 5 Liu Y, Ngu A, Yeng L. QoS Computation and Policing in Dynamic Web Service Selection. In: Proceedings of WWW, 2004
- 6 Vu Le-Hung, Hauswirth M, Aberer K. Towards P2P-based Semantic Web Service Discovery with QoS Support. In: Proceeding of Workshop on Business Processes and Services (BPS), Nancy, France, 2005
- 7 Vu Le-Hung, Hauswirth M, Aberer K. QoS-Based Service Selection and Ranking with Trust and Reputation Management. OTM Conferences, 2005(1): 466~483
- 8 Day J, Deters R. Selecting the Best Web Service. In: the 14th Annual IBM Centers for Advanced Studies Conf., 2004
- 9 Paolucci M, Kawamura T, Payne T, Sycara K. Semantic matching of Web services capabilities. In: Proceedings of the 1st International Semantic Web Conference (ISWC)s, 2002

(上接第 94 页)

4.3 请求信用值的消息转发分析

EigenTrust 在计算全局信用值的时候需要全局迭代, 严重的情况下将产生洪泛。而 CSBTrust 只在可信邻居节点范围内请求信用值, 因此它的消息转发数是很小的。从图 3 中可知随着系统总节点数的增加, EigenTrust 的信用消息转发数将成指数级增长, 很难在大规模网络实施。而 CSBTrust 模型的消息数基本保持在小范围内变化, 因此, 开销小, 具有很好的可扩展性。

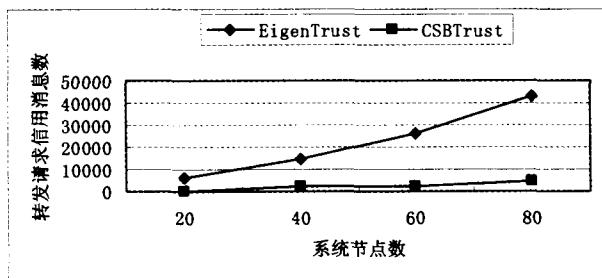


图 3 两种模型的消息转发数比较

**结论与进一步工作** 本文在已有模型的基础上构建了一个对等网络信用模型, 该模型利用节点的内容相似性来评价节点信用度, 并根据交易经验自适应动态调整邻居节点, 这种方式能充分地利用可信节点的推荐来选择下载服务。实验表

明该模型与同类模型相比, 有效提高了下载服务的质量, 并且具有良好的扩展性和较低的运行开销。

另外, 本文主要从请求服务的节点角度出发评价节点信用, 实际上提供服务的节点也需要对请求者进行信用评估, 抵制 free-riding 行为, 鼓励用户提供优质服务。因此, 以此为出发点进一步完善整个信用模型将是作者接下去的主要研究目标。

参考文献

- 1 Kazaa. http://www.kazaa.com
- 2 Liang J, Kumar R, Xi Y, Ross K. Pollution in P2P file sharing systems. In: Proceedings of IEEE INFOCOM 2005. IEEE Press, 2005, 2: 1174~1185
- 3 Xiong L, Liu L. PeerTrust: Supporting reputation-based trust for peer-to-peer electronic communities. IEEE Transactions on Knowledge and Data Engineering, 2004, 16(7): 843~857
- 4 Cornelli F. Choosing reputable servants in a P2P network. In: Lassner D, ed. Proc. of the 11th Int'l World Wide Web Conf. Hawaii: ACM Press, 2002. 441~449
- 5 窦文, 王怀民, 贾焰, 等. 构造基于推荐 Peer-to-Peer 环境下的 Trust 模型. 软件学报, 2004, 15(4): 571~583
- 6 Kamvar S D, Schlosser M T. EigenRep: Reputation management in P2P networks. In: Lawrence S, ed. Proc. of the 12th Int'l World Wide Web Conf. Budapest: ACM Press, 123~134
- 7 袁巍, 李津生, 洪佩琳. 一种 P2P 网络分布式信任模型及仿真. 系统仿真学报, 2006, 18(4): 938~942
- 8 Query Cycle. http://p2p.stanford.edu/www/qcsim.htm