

# 基于Chord网的语义Web Service发现\*

吴万明, 吴毅坚, 赵文耘

(复旦大学计算机科学与工程系 软件工程实验室, 上海 200433)

E-mail : 052021142@fudan.edu.cn, wuyijian@fudan.edu.cn, wyzhao@fudan.edu.cn

**摘要:** 随着 Web Service 数目的急剧增长和对实时性需求的增加, 集中式的服务发现机制已难以满足需求。分布式的服务发现机制由于其没有单点失败性及扩展性强的特点, 越来越受到关注。提出一个实用、高效、可扩展的新型语义 Web Service 发现模型, 将集中式的服务注册和服务发现机制分散到分布式网络结构中的各个注册节点, 并引入本体概念组和服务特征向量概念。首先对服务语义本体分组, 然后将服务描述和服务查询映射为服务特征向量进行匹配, 从而提高了服务发现的效率和准确性。

**关键词:** 语义 web service; web service 发现; peer-to-peer; chord; 本体

**中图分类号:** TP311

**文献标识码:** A

## Chord-based Semantic Web Service Discovery

WU Wan-ming, WU Yi-jian, ZHAO Wen-yun

(Department of Computer Science and Engineering, Fudan University, Shanghai 200433, China)

**Abstract:** Current web service discovery mechanism requires replication of all web service publications all registries. Large growth in number of Web services as well as the growth in the number of registries makes this replication impractical. Distributed web services discovery mechanism is put forward for its distribution and scalability. We introduced an approach for semantic discovery of web services in Chord-based registries. We use the structured peer to peer overlays as the service repository network, and we distribute semantic service advertisements among available registries such that it is possible to quickly identify the repositories containing the probable matching service. After advertising we propose the conception of Concept Group and Service Characteristic Vector to implement a scalable, efficient and reliable service discovery framework.

**Keywords:** semantic web service; web service discovery; peer-to-peer; chord; ontology

## 1 引言

典型的 Web Service 采用了面向服务的体系结构，这种体系结构包含了三个重要角色：服务请求者(Requestor)、服务提供者(Provider)和注册中心(Registry)。传统的 Web 服务发现机制使用集中的服务注册中心来存储和现服务描述，然而随着 Web Service 数目的增长和对实时性需求的增加，集中式的服务发现机制难以满足实际需求，因为集中式的服务发现机制有着一下几点难以克服的固有缺点，这包括(1)高维护成本，(2)单点失败性，(3)可扩展性低。为了克服这些缺点，基于 P2P 的分布式服务发现结构应运而生。

本文提出了一个实用、高效、可扩展的新型语义 Web Service 发现模型，它基于 Chord[1] P2P 网络结构以及本体论(Ontology)[2]将集中式的服务注册和服务发现机制分散到分布式网络结构中的各个注册节点。我们使用 OWL-S[3]语言描述 Web Service，服务语义特征(输入、输出等)主要通过 Service Profile 描述。通过将复杂的本体概念散列到简单的数值关键字，我们可以快速而精确的发现与用户的请求最匹配的服务，即提高服务的查准率。同时我们对服务的特征本体概念进行分类，根据本体概念之间的语义相似性将不同的本体概念分类成不同的本体概念组(Concept Group)，例如在本体结构图中的父类、子类可以归为同一个概念组，从而有助于提高语义服务发现的查全率。

## 2 相关工作

当前对 P2P 的研究分为两大类，一类是非结构化的网络，以 FreeNet、Gnutella 为代表，它们的缺点是对可扩展性缺乏有效的支持，当节点的数量增大时，系统的性能会急剧下降。另一类采用结构化的底层结构，以 Chord[1]、CAN。这类结构对可扩展性和稳定性提供了可靠的保证，本文即选用了这类结构中的一种：Chord。

本文模型采用结构化的 P2P 网络，引入了本体概念组和服务特征向量的概念，采用相容散列的路由算法检索注册服务，提高了服务发现的效率；同时网络中的每个注册节点是对等、自适应、自管理的，没有严格的依赖关系，因而具有良好的可扩展性。

## 3 Chord 网络模型

Chord[1]分布式查找服务提出了一种散列函数的快速分布式计算：给定一个关键字(key)，将关键字映射到系统中可靠的节点上。如果系统中每一个数据都指定了关键字，

那么数据分布式查找问题就容易解决了。Chord 采用了相容散列(Consistent Hashing)的一种变体计算方法为节点分配关键字。相容散列有几个很好的特点，首先是散列函数可以做到负载均衡。另外，当第 N 个加入或者离开系统时，只有  $1/N$  的关键字需要移动到其它位置。Chord 改善了相容散列的可扩展性。在 Chord 中，节点并不需要知道所有其它节点信息，每个 Chord 节点只需要知道其它节点的少量路由信息。在 N 个节点组成的系统中，每个节点只需维护其它  $O(\log N)$  个节点的信息。图 1a 和图 1b 分别给出了 Chord 的数据组织结构和 Chord 的路由查找方式。

## 4 基于 Chord 的语义 Web Service 发现模型

图 2 给出了一个在分布式环境下的 Web 服务发现模型框架。服务提供者将服务发布到 P2P 网络中的节点上。在 Chord 网络中，即是：(1)服务提供者将服务发布在自身所在的节点上(发布)；(2)服务请求者向 P2P 网络中的某个节点发出服务请求(请求)；(3)P2P 网络根据其路由算法找到合适的服务所在的节点(路由请求)；(4)将结果返回给服务请求者；(5)服务请求者选择其中一个满足条件的服务，并访问该服务的提供者，调用服务。

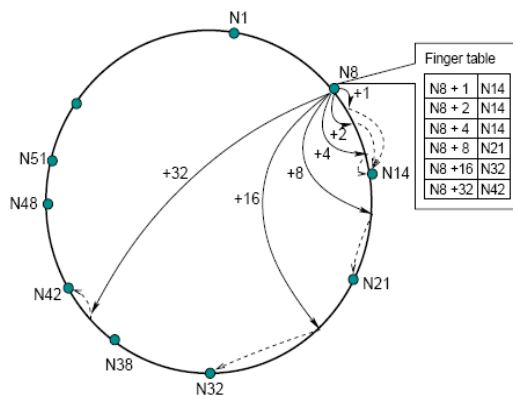


图 1a Chord 数据组织结构

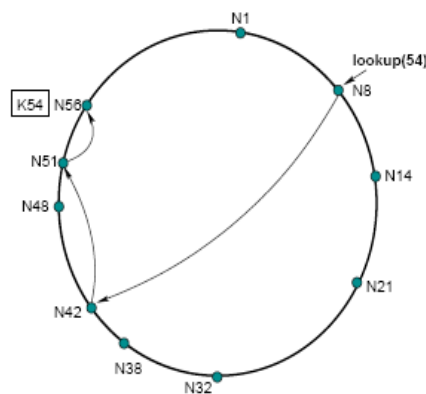


图 1b Chord 路由查找方式

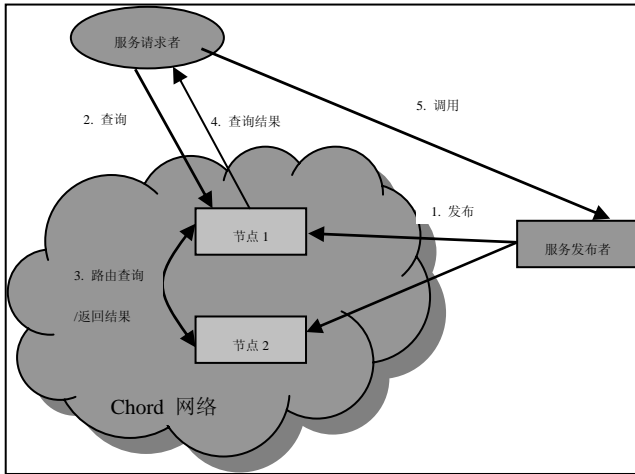


图2 分布式环境（Chord网络）Web服务发现模型框架

## 5 本体及语义 Web Service 发现

### 5.1 Web Service 的语义描述和服务特征向量

在 Chord 每个节点上所存储的 Web Service 使用 WSDL 及 OWL-s 协议来描述。

通过 WSDL 和 OWL-S 可以描述带有丰富语义信息的 Web 服务。为了高效地发布和查询这些带有语义信息的 Web 服务，我们引入了服务特征向量。服务的语义描述被映射成一个一维的向量。服务的发布和服务的查询都采用这个向量描述。服务发布时，我们使用服务特征向量描述一个具体的服务；与此相似，在服务发现过程中，服务请求者的请求(query)被转换成相对应的服务特征向量，与网络中已发布的服务特征向量进行匹配以获得合适的服务。

### 5.2 概念组和服务特征向量的计算

首先我们将描述服务的功能性信息，例如：输入(input)、输出(output)的本体概念进行分类，形成本体概念组(Concept Group)，如图3。本体概念组定义为相似语义的本体概念的集合，其中概念组之间的语义相似性由本体概念在本体结构图中的距离以及它们共同属性的数量多少来决定的，这部分工作在[5]中有具体的阐述。我们定义概念组之间具有大小可比性，其比较规则如下：

定义 1: 如果概念组  $CG_x$  大于概念组  $CG_y$ ，那么它必须要满足下列两个条件之一：

1. 概念组  $CG_x$  在本体结构图中的位置高于  $CG_y$
2. 概念组  $CG_x$  与  $CG_y$  在本体结构图中位于同一层，但是  $CG_x$  在图中的位置位于  $CG_y$  的左边。

在这个基础上我们将服务的语义描述(包括服务的发布和服务的查询)中的功能信息：输入(inputs)，输出(outputs)映射到它们相对应的概念组中。根据[6]将本体概念通过散

列的方法映射到数值可以有效地提高语义推理的能力，因此在本文中我们使用数值关键字(key)来代表服务的功能属性的本体概念。

我们通过 4 个步骤来完成使用 WSDL、OWL-S 描述的 Web Service 到服务特征向量的转变。

- (1) 使用 Bloom Key[4]散列算法来确定一组概念是否在语义上相似。
- (2) 对于每一个输入  $I_i$  或者输出  $O_i$ ，首先找到它所从属的概念组  $CG_i$ 。
- (3) 因为服务输入、输出参数的顺序对描述服务的功能属性没有影响，因而根据定义 1 我们将服务的输入、输出参数进行从大到小的排序。
- (4) 对于排序好的一个概念组  $CG_i$ ，使用 Bloom Key 散列算法生成每个  $CG_i$  所对应的关键字  $k_i$ ，所有  $k_i$  的组合即构成该服务的特征向量。

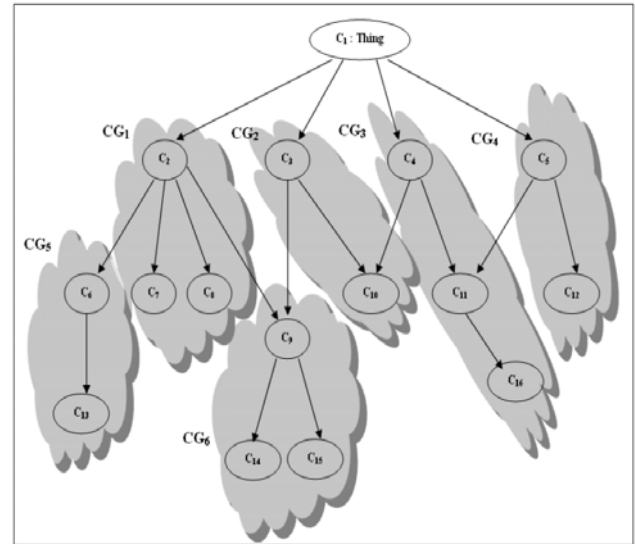


图3 本体与本体概念组

### 5.3 服务的发布

在本文提出的服务发现模型中，P2P 网络中的每个节点都接受服务注册，同时也接受服务查询请求。模型处理服务发布请求的步骤如下：

1. 对于服务  $S$ ，使用 Bloom 方法获得其对应服务特征向量  $V = \{k_1, k_2, \dots, k_p, d, k_{p+1}, \dots, k_q\}$ ， $k_i$  为服务的输入输出根据 Bloom 方法散列获得的 key， $d$  是输入与输出的分隔符。
2. 使用散列算法  $H$  将向量  $V$  整个映射到数值  $K$ ，可以选择合适的算法  $H$  使得  $K$  包含有向量  $V$  中每一个元素的特征。
3. 根据前文提到的 Chord 路由查询算法，查询出关键值  $K$  对应的节点  $P_i$ 。
4. 将服务  $S$  发布在节点  $P_i$  上。

## 5.4 服务的发现

模型处理服务发现请求的步骤如下:

1. 对于每个服务特征向量 $V=\{k_1, k_2, \dots, k_p, d, k_1 \dots k_q\}$ 。
2. 通过散列算法  $H$  将向量  $V$  整个映射到数值  $K$ 。
3. 根据 Chord 路由查询算法, 找到包含有关键值  $K$  的节点  $P_i$ 。
4. 节点 $P_i$ 中即保存着服务特征向量为 $V$ 的服务。

通过上述方法, 我们可以将用户提交的服务查询快速的匹配到某个节点上已注册的服务。由于使用了Bloom方法, 确定一个本体概念从属于哪个本体概念组这一步骤变得快速有效, 所以计算一个服务的特征向量这个过程也耗时很少。然而由于每个服务特征向量的元素个数可能不一样, 元素本身可能不一样, 匹配两个服务特征向量就成了本文所要解决的问题。为了解决这个问题, 我们考虑了两种不同的方法。第一, 将向量 $V=\{k_1, k_2, \dots, d, \dots, k_n\}$ 中的所有元素视为一个整体, 作为Chord网的搜索关键字。通过P2P路由算法检索到该关键字所在的节点, 则所要查找的服务就在该节点上。第二, 将向量 $V$ 中的每个元素单独作为一个关键字, 利用Chord路由查询算法检索到一系列的节点, 然后取这些节点的交集获得最匹配的节点。出于提高服务发现效率的考虑, 我们使用第一种方法, 因为服务之间的语义描述接口可能会千差万别, 服务的参数也可能会非常多, 每个参数匹配到的节点数量可能会很多, 多个参数检索到的节点数将是一个庞大的数目, 这么大量数的节点取交集是一个很大的工作量, 会大大降低服务发现的效率。

## 6 总结与未来工作

本文提出了一个基于 Chord 网络以及本体论的新型语义 Web Service 发现模型。模型中, (1)使用 OWL-S 语言描述 Web Service, Service Profile 描述服务语义特征。(2)通过对本体概念进行分组, 将服务映射到特定的服务特征向量。(3)根据 Chord 结构, 将服务发布到对应的节点上。(4)服务请求时: 根据 Chord 路由算法, 在网络中进行服务特征向量匹配, 发现匹配请求的节点。(5)请求者访问匹配节点, 调用服务。通过本模型, 我们实现了一个 P2P 分布式网络环境下的服务发现方法, 并且大大提高了服务发现的查准率和查全率。

在进一步的工作中, 我们将研究在不降低服务发现效率的前提下, 如何使得一个本体概念可以同时映射到多个本体概念组。

## 参考文献:

[1] Stocia R Morries, D Karger, *et al.* Chord: A Scalable

Peer-to-Peer Lookup Service for Internet Applications [J]. ACM SIGCOMM, 2001, 11(1):17-32.

[2] Gandon F. Ontology Engineering: A survey and a Return on Experience [J]. Institut de Recherche en Informatique et Automatique (INRIA), 2002, 11(2):964-971.

[3] D Martin, M Burstein, *et al.* OWL-S: Semantic Markup for Web Service[EB/OL].

<http://www.daml.org/services/owl-s/1.0/>, 2004.

[4] B H Bloom. Space/Time trade-offs in hash coding with allowable errors [J]. ACM, 1970, 13(7):422-426.

[5] S Castano, A Ferrara, S Montanelli, G Racca. Matching techniques for resource discovery in distributed systems using heterogeneous ontology descriptions [A]. S Latifi. ITCC'04: Proceedings of the International Conference on Information Technology: Coding and Computing[C]. Washington, DC: IEEE Computer Society, 2004.360-366.

[6] I Constantinescu, B Faltings. Efficient matching and directory services [A]. Ning Zhong. WI'03 Proceedings of IEEE/WIC International Conference on Web Intelligence[C]. Washington, DC: IEEE Computer Society, 2004.75-81.

## 作者简介:

吴万明 男, 江苏省扬州市人. 复旦大学计算机科学与工程系硕士. Email: [052021142@fudan.edu.cn](mailto:052021142@fudan.edu.cn).

吴毅坚 男, 上海人. 博士. 主要研究方向为软件构件技术, 面向服务架构和网格计算. Email: [wuyijian@fudan.edu.cn](mailto:wuyijian@fudan.edu.cn).

赵文耘 男, 江苏常熟人. 教授, 博士生导师. 主要研究方向为软件工程. Email: [wyzhao@fudan.edu.cn](mailto:wyzhao@fudan.edu.cn).