

内容寻址网络的 P2P 语义Web服务组合系统架构

朱正东¹, 胡亚红², 伍卫国¹, 王勇³, 李增智¹

(1. 西安交通大学电子与信息工程学院, 710049, 西安; 2. 浙江工业大学信息工程学院, 310032, 杭州;
3. 开封广播电视大学, 475100, 河南开封)

摘要: 为了提高语义 Web 服务组合系统的扩展性、可靠性和稳定性,提出了一种基于 P2P 的语义 Web 服务(P2PSWS)组合系统架构. 该架构结合了集中式和分布式结构的优点,将统一描述、发现和集成协议的功能分散于本地 Web 节点、组 Web 节点和公共 Web 节点;设计了一种基于内容寻址网络的 P2P 网络语义 Web 服务的定位机制,以保证每个共享服务按领域划分,按被所有节点所共享的公共 Web 节点来注册. 原型系统运行表明,该架构有助于克服传统公共 Web 服务结构上的单一节点失败问题,扩展了 P2P 系统的能力,有效地实现了基于本体的语义 Web 服务的组合.

关键词: 语义 Web 服务;服务组合;内容寻址网络

中图分类号: TP311 **文献标志码:** A **文章编号:** 0253-987X(2010)02-0006-05

CAN-Based P2P Semantic Web Services Composite Architecture

ZHU Zhengdong¹, HU Yahong², WU Weiguo¹, WANG Yong³, LI Zengzhi¹

(1. School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China;
2. College of Information Engineering, Zhejiang University of Technology, Hangzhou 310032, China;
3. Radio and TV University of Kaifeng, Kaifeng, Henan 475100, China)

Abstract: A P2P-based semantic Web services (P2PSWS) composition system architecture is presented to improve the scalability, reliability and stability of semantic Web service composition systems. The architecture combines the advantages of centralized structure and distributed structure, and distributes the function of UDDI to the local Web node, group Web nodes and public Web nodes. A location mechanism of semantic Web services is designed based on CAN (content-addressable network) P2P networks to ensure that each service is registered in one node of the public Web. The registration services of public Web nodes are divided by areas and shared by all nodes. The running of prototype system shows that the proposed structure helps to overcome the problem "the failure of a single node" on the traditional structures of public Web services and expands the capacity of P2P systems, and achieves an efficient ontology-based semantic Web services composition.

Keywords: semantic Web service; service composition; content-addressable network

近年来, P2P 技术加强了分布式计算结构的可靠性和扩展性^[1], 由此 Web 服务便成为分布式环境中服务共享架构的基础. 语义 Web 是一种新的结构, 它通过智能标记语言、Web 本体语言(OWL)

增强了形式语义的内容, 使之更适合于机器的自动推理和消费, 支持服务的自动发现、组合、执行和互操作.

近来, 在内容寻址网络(CAN)中给出了一种组

需求时在知识库中使用语义匹配方法寻找相应的 Web 服务模板,或者按照模板、根据语义匹配方法动态地发现子 Web 服务,并进行绑定;组合引擎能够在未找到组合 Web 服务模板时,基于需求及语义匹配方法进行 Web 服务的自动组合,将得到的组合 Web 服务经验证后作为模板保存于知识库之中;执行引擎将模板转换为 BPEL4WS 的具体工作流程,通过 Web 服务调用框架(WSIF)协议管理和执行与用户需求一致的 Web 服务。

Web 服务提供者将在注册库发布 Web 服务,同时为知识库提供 Web 服务模板,并从 Web 服务的 OWL-S 相关描述中提取 Web 服务关系本体。Web 服务 QoS 发布是为提供者提供输入 Web 服务 QoS 矩阵值的一个用户接口,Web 服务 QoS 反馈是为执行引擎对使用的 Web 服务进行评估而设置的应用程序接口。

2 Web 服务的处理

搜索引擎主要完成 Web 服务的发布与发现,组合引擎主要完成 Web 服务的组合和验证,执行引擎主要完成 Web 服务的管理和执行,所以使用 OWL-S 描述的 Web 服务处理模块包括服务发布、服务发现、服务组合、服务验证、服务调度和服务执行等。由于篇幅关系,本文重点描述定位和发现方法,其他内容参见作者与同行的相关研究^[7]。

2.1 Web 服务的定位方法

服务的发布首先要确定服务发布的位置,所以 Web 服务的定位方法便成为服务发布的关键。本系统使用 CAN 的 P2P 基础结构,提供分布式哈希表,利用关键词和术语向量查询相关对象。在笛卡尔空间,对象是空间中一个代表查询文件的节点,对于公共 Web 服务它就是 CWP。每一个 CWP 负责空间中的一部分区间,区间中的每个 Web 服务将在 CWP 中进行注册,注册信息包括 Web 服务的位置,该位置可用相应的 IP 地址表示。使用 VSM 和 LSI 可将每份文件与文件术语矩阵对应的术语向量进行映射。VSM 是信息检索领域的经典统计算法,它将描述文档的多个特征词以向量的形式来表示,向量中的每个分量对应了文档的不同特征。本系统将 Web 服务作为文档,在 Web 服务的发布和检索中使用 VSM 是将 Web 服务描述中的关键词和术语作为特征词来处理,特征词可区分 Web 服务的所属领域。下面用一个二维空间的例子来说明 Web 服务在 P2P 网络的 CWP 中的定位问题。

将 Web 服务描述经处理后得到的关键词作为横坐标,输入输出参数作为 Web 服务术语的纵坐标,不能用来区分服务领域的其他内容将不在考虑之中,如服务的动作与质量指标。

假设整个 P2P 网络中包含的公共 Web 服务的集合 S 由 n 个 Web 服务构成,即 $S = \{s_1, s_2, \dots, s_n\}$ 。 S 中所有 Web 服务包含的关键词构成集合 $C = \{c_1, c_2, \dots, c_m\}$, S 中包含的术语构成集合 $D = \{d_1, d_2, \dots, d_k\}$ 。如果 $s_i \in S$, 则 s_i 的坐标为 (x, y) , 其中: $x = x_{i1}x_{i2}x_{i3} \dots x_{ij} \dots x_{im}$, $x_{ij} = 1$ 表示 s_i 中包含第 j 个关键词 c_j , 否则 $x_{ij} = 0$; $y = y_{i1}y_{i2}y_{i3} \dots y_{ij} \dots y_{ik}$, $y_{ij} = 1$ 表示 s_i 包含第 j 个术语 d_j , 否则 $y_{ij} = 0$ 。

假设上述例子中 m, k 均为 5, 结果如图 3 所示。每个区间可表示不同领域的服务,并对应一个 CWP 的 IP 地址,该 IP 地址的站点注册了一个领域的所有 Web 服务。图 3 中, $A \sim F$ 是节点,分别负责管理某一特定领域的空间(矩形部分),其中已发布的 Web 服务位于 (x, y) , 应该在节点 A 登记, 因为它的坐标位于 A 区域。

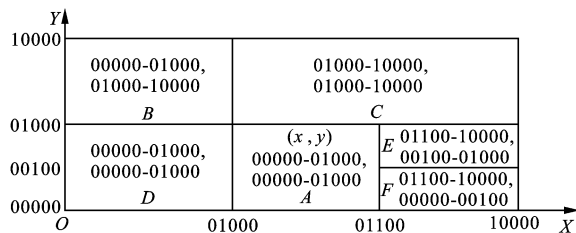


图3 Web 服务的定位

2.2 Web 服务的发现方法

下面以上述二维空间的例子来阐述 Web 服务的发现方法。首先定义关键词及术语的词频向量 $w_i = \{n_{i1}, n_{i2}, \dots, n_{ij}, \dots, n_{il}\}$, 其中 $l = m + k$, n_{ij} 是 s_i 包含的第 j 个关键词或术语的数量,可采用 TFIDF (Term Frequency Inverse Document Frequency)^[8] 来计算。本文用 $n_{ij} = f_{ij} \ln(N/M_j + 0.5)$ 计算关键词和术语的权重,推断关键词和术语的重要性,其中 f_{ij} 是第 j 个关键词或术语在 s_i 中出现的次数, N 是 Web 服务集中出现 Web 服务的总数, M_j 是包含第 j 个关键词或术语的 Web 服务的数量。

在用定位方法确定了 Web 服务注册的位置(节点 CWP)后,提取关键词术语向量 $I = (I_1, I_2, \dots, I_j, \dots, I_l)$, 其中 $l = m + k$ 。如果用户输入的关键词和术语中包含第 j 个关键词或术语,则 $I_j = 1$, 否则 $I_j = 0$ 。将用户的关键词术语向量与 Web 服务中的 $w_i = \{n_{i1}, n_{i2}, \dots, n_{il}\}$ 进行点乘,可得到 Web 服务的

相似度 $r_l = \sum n_{lk} I_k (k = 1, l)$, 可见 r_l 越高, 与 Web 服务的关联度越高, 就越符合用户的要求. 事实上, 本系统未设立任何维数和维序限制, 通常情况下用户可以根据需求建立多维空间.

3 P2PSWS 原型系统

根据上述架构实现了 P2PSWS 原型系统——一种基于 P2P 的语义 Web 服务组合原型系统. 该系统采用 JXTA 作为基本的 P2P 系统平台^[5], 并在局域网中设置了 3 个组, 分别为餐饮组、书店组和混合组, 每个组由 3 台 PC 机组成, 其中 2 台为 LWP, 1 台为 GWP. 另外, 还有 2 台 CWP, 分别用来注册餐饮 Web 服务和书店 Web 服务. LWP 使用了 Windows XP, GWP 与 CWP 使用了 Windows 2003 企业版服务器软件, 所有机器均安装了 P2PSWS 软件, 数据库使用了 SQL Server 2000, 本体用 OWL-S. 下面描述原型系统中服务发布与用户需求的处理过程, 实验使用文献^[9]中许多成熟的服务实例及开发的餐饮服务实例.

3.1 用户发布服务的过程

服务发布者包括 OWL-S 文件处理器、服务定位器、服务注册器, 如图 4 所示. 文件处理器从 OWL-S 文件中提取相关领域的本体、关键词和术语, 并将关键词和术语传送到服务定位器上; 定位器计算 Web 服务的坐标, 确定在 CAN 上发布 Web 服务的具体位置; 注册器对用于 Web 服务的位置管理节点进行注册, 该节点要维护本地服务的缓存, 存储发布的 Web 服务, 消除生命周期结束的 Web 服务.

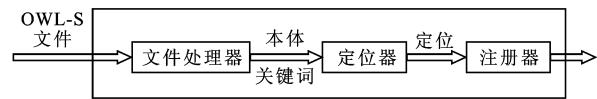


图 4 服务发布的过程

具体地说, 首先要处理包含服务信息的 Web 服务实例, 它的输入是 3 个 OWL-S 形式的文件: ServiceProfile、ServiceModel 和 ServiceGrounding. OWL-S 处理过程包括 OWL-S 功能模块的生成、OWL-S 在各个节点间的传输以及节点对 OWL-S 的解析. 功能模块的生成过程采用 Java2 OWL-S 工具, 传输利用了 P2P 系统的共享档案机制, 解析使用了 OWL-S4J 工具包.

3.2 用户需求处理过程

由于 Web 服务是通过服务广告发布的, 所以发

现 Web 服务就是寻找适当的广告, 并获得相应的 OWL-S 文件. 图 5 显示了在 P2PSWS 系统中发现 Web 服务广告、组合 Web 服务及其选择 Web 服务的算法流程.

当用户提出需求时, LWP 首先查询本地缓存中的服务广告, 如果没有查到, 将需求传送到 GWP 再进行查询, 若仍未查到, 则由 GWP 对 CWP 进行定位, 以发现匹配的服务广告; 否则, 将需求提交到本地 GWP 组合引擎进行服务组合. 在发现 Web 服务广告的同时, 节点将得到广告所指明的 OWL-S 文件, 同时绑定 Web 服务, 并通过使用 SOAP 要求提供、执行 Web 服务.

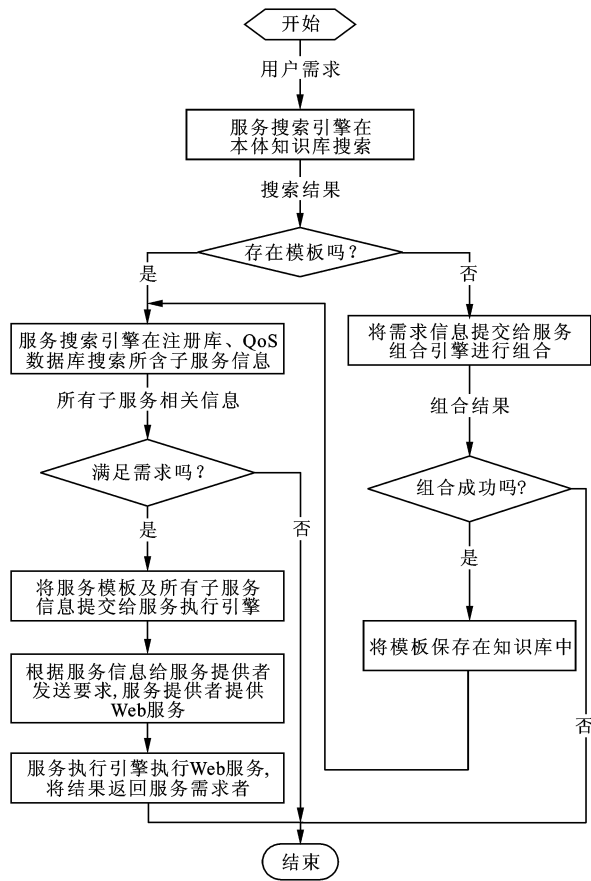


图 5 用户需求处理流程

综上所述, 本文所提架构的优点是: 在使用 GWP 和 CWP 注册新服务时具有灵活性, 在站点和服务数量很多时, 可减少 CWP 的负载; 要求提供服务的需求不会被淹没在整个网络系统之中; 结构中任何一个 GWP 和 CWP 出现故障时, 系统都可以通过其他节点恢复 Web 服务的相关数据.

4 结论及未来工作

本文提出的基于 P2P 技术的 P2PSWS 体系结构充分利用了 P2P 网络及其分布式架构的优势,避免了 UDDI 机制存在的单一节点失败问题,增强了系统的可靠性. 基于 CAN 的基础结构,分发语义可服务于 P2P 结构的各节点,同时为提高语义 Web 服务组合系统的扩展性和稳定性,建立了一个分层次的语义 Web 服务系统. 在原型系统上实现的基于向量空间模型和语义索引的服务发布、发现和组合方法,能更有效地寻找服务资源,提高了不同 P2P 系统之间的互操作性.

参考文献:

- [1] 马建刚,黄涛,汪锦岭,等. 面向大规模分布式计算发布订阅系统核心技术 [J]. 软件学报, 2006, 17(1):134-147.
MA Jiangang, HUANG Tao, WANG Jinling, et al. Underlying techniques for large-scale distributed computing oriented publish/subscribe system [J]. Journal of Software, 2006, 17(1):134-147.
- [2] COHEN E, FIAT A, KAPLAN H. Associative search in peer to peer networks: harnessing latent semantics, computer networks [J]. The International Journal of Computer and Telecommunications Networking, 2007, 51(8):1861-1881.
- [3] HAKEM B, GEERT D. Comparing chord, CAN and pastry overlay networks for resistance to DoS attacks [C] // Proceedings 2008 3rd International Conference on Risks and Security of Internet and Systems. Piscataway, NJ, USA: IEEE, 2008:261-266.
- [4] VERMA K. METEOR-CS WSDI and 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.
- [5] FATOS X, THANASIS D, LEONARD B. Efficient peer selection in P2P JXTA-based platforms [C] // Proceedings of 22nd International Conference on Ad-

vanced Information Networking and Applications. Piscataway, NJ, USA: IEEE, 2008:1013-1020.

- [6] ALDA S. Peer group-based dependency management in service-oriented peer-to-peer architectures [M]. Berlin, Germany: Springer, 2007:195-209.
- [7] ZHU Zhengdong, DONG Xuehan, HU Yahong, et al. A minimum coverage dynamic adaptive method for concurrent Web service composition [C] // Proceedings of IEEE Asia Pacific Services Computing Conference. Piscataway, NJ, USA: IEEE, 2008:1083-1088.
- [8] QU Shouning, WANG Sujuan, ZOU Yan. Improvement of text feature selection method based on TFIDF [C] // Proceedings of 2008 International Seminar on Future Information Technology and Management Engineering. Piscataway, NJ, USA: IEEE, 2008:79-81.
- [9] Web-Ontology Working Group. OWL-S 1.0 release: examples [EB/OL]. [2009-01-10]. <http://www.daml.org/services/owl-s/1.0/examples.html>.

[本刊相关文献链接]

基于扩展投影时序逻辑的组合 Web 服务描述与验证. 2007,41(10):1155-1159.
可用性语义 Web 服务的通用发现机制. 2008,(6):659-663.
一种基于最小覆盖的复杂 Web 服务组合方法. 2008,42(8):945-949.
广义随机 Petri 网下的组合 Web 服务建模与评价. 2008,42(8):967-971.
一种编辑距离算法及其在网页搜索中的应用. 2008,42(12):1450-1454.
面向投影时序逻辑的 Web 服务模型检测. 2009,43(4):39-43.
语义 Web 中的概念等级匹配算法. 2009,43(4):49-51.
面向 Pareto 最优遗传算法的服务组合方法. 2009, 43(12):50-54.

(编辑 苗凌)