

web 服务发现综述

张通¹, 王丹志², 鄂海红²

(1. 北京邮电大学计算机学院 PCN&CAD 实验室, 北京 100876;

2. 北京邮电大学计算机学院, 北京 100876)

摘要: web 服务基本流程包括: 服务提供者将其服务发布到服务代理的一个目录上; 当服务请求者需要调用该服务时, 首先利用服务代理提供的目录去搜索服务的信息; 然后服务请求者根据这些返回的相关信息去调用服务提供者发布的服务。最新关于 web 服务发现机制被定义为“通过对 web 服务的处理描述进行定位, 以满足特定的功能需求”。服务提供者的发现机制首次被描述为匹配过程。也就是服务请求者通过中间代理寻求合适服务提供者的过程。进行服务发现目的就是为了消除在可用 web 服务之间的异质性, 因此可以将 web 服务发现定义为一套满足 web 服务发布和 web 服务请求, 能满足 web 服务请求, 并保证 web 服务可用性和可信性的机制。

关键词: 计算机技术; web 服务; 服务发现

中图分类号: TP391.7

Overview of web service discovery

Zhang Tong¹, Wang Danzhi², E Haihong²

(1. PCN & CAD Lab, Beijing University of Posts and Telecommunications, Beijing 100876;

2. Beijing University of Posts and Telecommunications, School of Computer Science, Beijing 100876)

Abstract: The web service basic processes including: service providers to publish their services to a directory service agent; service requester need to call the service provided by the service agent directory search service; service requestor then according to the return of the relevant information to call the service by the service providers. Latest web service discovery mechanism is defined as "processing web services description positioned to meet the specific functional requirements. The service provider discovery mechanism is described for the first time for the matching process. Service requester through an intermediate agent to seek appropriate service provider process. Service discovery purpose is to eliminate the heterogeneity between the available web services, so the web service is defined as a set to meet web publishing and web service request, to meet the web service request, and to ensure that the web service availability and letter of the mechanism.

Keywords: Computer Technology; Web Services; Service Discovery

0 引言

随着核心 web 服务标准(WSDL、SOAP)及补充协议规范逐渐被采纳和实现, web 服务成为异构环境和异构技术中的网络化应用操作集成标准, 面向服务的架构因为其可变性和动态性得到了广泛的关注和应用。

作者简介: 张通, (1986-), 男, 硕士研究生, 服务科学与工程、宽带无线移动系统与互联网。

通信联系人: 王丹志, 男, 副教授, 无线 MESH 网络、无线宽带通信技术、移动增值业务。 E-mail: danzhiwang@yahoo.com.cn

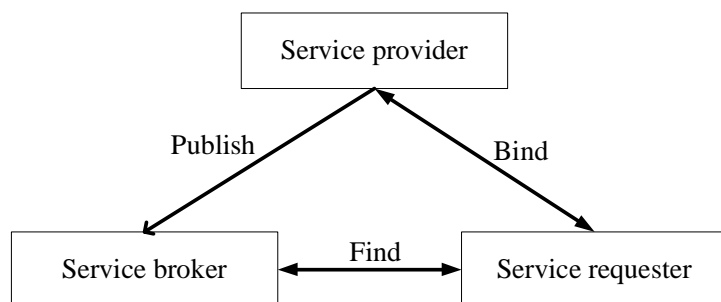


图 1 web 服务基本架构

Fig. 1 web services basic architecture

从图 1 中 web 服务基本架构可以看出，其中包括三个基本操作和三个参与者。服务提供者将其服务发布到服务代理的一个目录上；当服务请求者需要调用该服务时，首先利用服务代理提供的目录去搜索服务的信息；然后服务请求者根据这些返回的相关信息去调用服务提供者发布的服务。

服务提供者的发现机制首次被描述为匹配过程。也就是服务请求者通过中间代理寻求合适服务提供者的过程^[1]。最新关于 web 服务发现机制的定义更加宽泛，定义为“通过对 web 服务的处理描述进行定位，以满足特定的功能需求”^[2]。

进行服务发现目的是为了消除在可用 web 服务之间的异质性，这些不同的异质性包括了^[3]：

- a. 技术异构性（不同的平台或数据格式）
- b. 拓扑异构性（服务中特定的领域和概念）
- c. 实用异构性（特定领域过程的不同发展和特定领域任务的不同概念）

这里，可以将 web 服务发现定义为一套满足 web 服务发布和 web 服务请求，能满足 web 服务请求，并保证 web 服务可用性和可信性的机制。本论文将从 web 服务的保存维护、搜索和可用性保证三个方面来论述 web 服务发现相关的研究与进展，同时对相关实验室方法和评价进行介绍，方便 web 服务发现的持续研究。

1 Web 服务的保护维护

Web 服务发现机制是保障 web 服务满足大量广泛应用、充分得到利用的基础。Web 服务的诞生，带来了大规模的复用，软件逐步呈现分布化和网络化的趋势。将软件作为组建进行复用的思想已经产生了近 40 年^[4]，而且证实了由人工对这些内容进行管理注定是要失败的^[5,6,7]，主要的原因如下：

- a. 内容的质量不断降低而变得不可用；
- b. 由于内容存储容量巨大而导致不易管理和维护。

目前进行 web 服务保存和维护方便发现的主要两个方式就是：注册和非注册两种。区分这两种方式的不同主要看的是：注册的方式就是通过特定的协议和描述对服务进行保存和维护，而非注册方式则是制定特定的服务方式。

1.1 注册方式

在 web 服务的三大协议中，UDDI 协议被用来进行服务的发布和发现^[8]，但 UDDI 有两个主要的缺陷：单点失效导致的缺乏可扩展性；按固定分类进行搜索带来的低精确性。

针对第一个缺陷，解决的主要思想就是让公开可用的 UDDI 节点组成一个服务，在外看来就像是一个节点在工作，可以把这些 UDDI 节点称为 UDDI 云或联合 UDDI^[9]。Tewari 等

人^[10]对现有的 SOA 架构进行了扩展, 通过多个 UBR 节点作为分布式存储进行服务的注册, 利用 IP 表对分布式节点进行管理, 克服 UDDI 注册服务器单点失败的问题。Wang 等人^[11]提出了利用 P2P 覆盖网解决传统的 UDDI 单点问题, 利用服务分类建立领域拓扑, 由领域节点组成分布式 Chord 网络, 通过关键字匹配进行领域节点的定位。Kashani 等人^[12]提出基于内容相似度构建 P2P 网络 QDN^[13]。解决单点登录问题的方法, 主要是采取 P2P 的思想, 让多个 UDDI 服务器进行协同工作。其中的研究热点有: 组织方式(集中式、完全分布式等)、节点组织(服务分类、内容相似度、功能相似度等)、节点的查找(关键字匹配、语义匹配等)。

针对固定分类给 web 服务浏览和搜索带来的不便, 目前有多种解决方案。利用标签进行直接和轻量的水平分类^[14,15], 但标签存在着非结构化(水平化)、信息垃圾和语义模糊以及语义丢失等问题^[16]; 利用数据挖掘或文本挖掘中的聚类算法进行分类^[17,18]。

1.2 非注册方式

对于非注册, 主要采取本地存储和网络爬取, 以及两种形式结合的方式进行服务的存储和索引。目前网络上出现了很多代码搜索引擎, 一般使用 Lucence 来实现索引机制。目前市场上最大的四个代码搜索引擎是 Koders.com/Krugle.com/merobase.com 和 Google.com, 截止 2011 年, 其对比表如下。

表 1 搜索引擎对比
Tab. 1 Search engines comparison

	Koders	Krugle	merobase	Google
Web 服务组件数目	~3M	~6M	~10M	~6M
支持语言数目	37	32	48	46

一般搜索引擎一方面是通过 web 服务发布者提供详细 web 服务信息, 另一方面通过网络爬虫进行 web 服务信息的爬取。如果是通过爬虫获取的相关 web 服务信息, 其中的问题包括了: 对 WSDL 文件的识别以及对 web 服务可用性的监测^[19]。

2 Web 服务的搜索

搜索的目的就是为了让用户的需求和可用的服务相匹配的过程, 其总的来说都是通过对 web 服务特征的提取和用户请求进行匹配。Web 服务特征的提取和 web 服务的分类都是基于 web 服务的相似性。

2.1 web 服务相似性

目前的分类有 UDDI 的固定分类、标签的水平分类, 以及各种聚类。进行非固定分类时, 都要对 web 服务进行特征抽取, 作为分类时的参数, 建立关联关系进行相似性比较。

基于 web 服务功能相似性。Khalid 等人通过从 WSDL 文档中抽取五个特定的属性形成最后的归一化相似度两, 使用 Quality Threshold 聚类算法进行聚类。Qi 等人^[20]将服务和操作都进行特征抽取, 而操作往往代表 web 服务的功能, 并通过两者的共现矩阵转换为二部图, 并利用 SVD 方法进行聚类^[21,22]。

Zhao 等人^[23]提出基于语义和结构相似性的服务关系自动发现算法, 语义相似度计算采用 cosine 相似度^[24,25], 结构匹配采用树结构, 也可基于图结构。

基于 web 服务消息或行为。Mealy 服务模型^[26,27]、BPEL^[28]和 WSCI 模型^[29]都是基于消息的模型; Roman 模型^[30]、PSL^[31]、CTR-S^[32]和 OWL-S^[33]都是基于事件或行为的模型。此

外 Shen 等人^[34]提出了消息和行为的混合模型,对 IOPR 模型进行拓展,将 web 服务的消息和行为进行关联,形成非确定有限自动机。

基于 web 服务的操作^[35,36]。将操作建模为“词袋”(term-bags),通过词袋之间的相似性进行发现。为了提高精确性,相似的词聚类为概念,这样用于操作的对比和匹配。但这样会有一些缺点:在 WSDL 文件中对操作和参数的定义有限;形成的特征维度较高;操作包含相似词语的 web 服务在功能上可能大相径庭。

2.2 web 服务的匹配

Web 服务之间、web 服务和用户请求之间的匹配主要有两种形式,一种是基于关键字匹配,另一种是基于语义的匹配。

关键字匹配。进行关键字匹配,一般都是在对 web 服务进行分类的基础之上进行。UDDI 中就提供基于关键字匹配的 web 服务搜索;[11]中在 P2P 覆盖网中进行领域节点搜索时,也是基于关键字匹配进行分类搜索的。基于关键字匹配的搜索,准确率和回调率都较低。一般都是讲基于关键字的匹配和其它匹配方式结合。

基于语义的匹配。在对 web 服务进行分类中,最常用的就是在建立 web 服务相似性模型的基础上,利用 term-bags 建立向量机,从而对 web 服务进行语义相似性计算,作为匹配的依据,包括利用 cosine 相似度、NGD 算法^[37]、利用 TF/IDF 建立权值矩阵、MatchMaker 等。此外,还有借助 WSIL、OWL-S 和语义标注 WSDL^[38]进行语义匹配。

目前随着语义 web 的兴起,通过构建语义网络便于 web 服务的发现正成为热点。Luo 等人^[39]通过 E-FCM 实现自动化本体描述构建 ALN 关联链接网络,将 web 服务进行组织为 C-ALN 网络,具备较短路径,便于快速服务发现和高效推荐。此外,p2p 覆盖网上还有一层语义网,覆盖网中领域节点的匹配是基于关键字的,而语义网中的匹配则是利用 Gnutalla 协议进行通信和基于一阶逻辑进行推理。

2.3 排序与推荐

为了便于用户选择合适的 web 服务,对搜索结果的展示和推荐,也关系到服务发现的精确性。

在搜索结果的展示上,一般会对搜索结果进行排序。目前进行服务导航可以根据服务全局重要性,可以通过 PageRank^[40]或 HITS^[41]计算。

在进行推荐方面,[20]中利用 web 服务组织性进行推荐。

3 Web 服务的可用性保证

为了保证搜索结果的 web 服务可用性,对于 web 服务可用性的保证主要考虑的有:可用性监测的触发时机(非搜索和搜索)、监测的位置(服务端、用户端)、测试数据(服务端提供数据、收集客户端运行数据)等。

[19]中,建立了内置执行引擎。该执行引擎将用户利用该引擎测试服务时的数据保存在数据库中作为测试数据,对服务的测试包括用户主动执行触发和后台周期性触发。在两种触发情况下发现服务不可用时,那么后台周期性监测的周期会减少,当测试次数达到一定数值,该 web 服务将被认为失效而被删除。

[10]中,通过 Validate WS 组件在搜索结果返回之前,对搜索结果进行可用性测试,保证 web 服务可用性。

4 相关实验方法和评价参数

4.1 相关实验方法

主要对目前的实验数据搜集和语义匹配两个方面进行了小结。

在实验数据收集方面。目前主要是在搜索引擎和 `bindingpoint/salcentral/xmethod/webservicex/webservicelist` 等 web 服务发布网站, 由于对 OWL-S 的缺乏, 一般在其上进行 WSDL 文件的下载; 另一方面是对 UDDI 注册中心进行访问, 获取其中的 WSDL 数据, 有三种方式: 使用基于 Java 的 SOAP API、使用普通的基于 Java 的 UDDI 客户端 API 和使用 JAXR(Java 为 XML 注册提供的 API)。

在进行语义匹配方面。包括取词工具 WVTool(Java), 以及利用 Porter Stemmer^[42]进行词语形态转换和词语预处理。

4.2 服务发现评价参数

Web 服务发现的评价。主要是两个参数: 准确率和回调率, 此外还有可扩展性等。

准确率是指 web 服务发现的结果中真实关联 web 服务所占的比率; 而回调率则是指 web 服务发现中关联服务的返回率。

假设针对特定的 web 服务发现请求, 相关的 web 服务集合用 `relevant` 表示, web 服务发现结果的 web 服务集合用 `retrieved` 表示, 那么准确率和回调率的表达式分别为:

$$\text{准确率} = \frac{(\text{relevant} \cap \text{retrived})}{\text{retrived}}, \quad \text{回调率} = \frac{(\text{relevant} \cap \text{retrived})}{\text{relevant}}。$$

5 结论

本论文从保存维护、搜索和可用性保证三个方面对目前 web 服务发现相关的研究进行了简单的归纳与小结, web 服务发现是服务组合和服务重用的基础, 也是 web 服务广泛使用的基础。目前对 web 服务发现的研究角度转变趋势有: 语义化和网络化。语义化让 web 服务更好的满足特定用户的需求, 而网络化则将 web 服务作为整体考虑, 提升 web 服务发现的效率和准确度, 同时为 web 服务组合提供基础性结构支持。

致谢

感谢宋美娜博士在我攻读硕士学位时的研究方向上的指正, 感谢鄂海红老师在论文写作中给予的意见, 感谢童俊杰博士和刘廉如博士给出的技术指导。

[参考文献] (References)

- [1] Decker K, Sycara K, Williamson M. Middle-Agents for the Internet[A]. Proc. 15th IJCAI[C]. Nagoya: Japan, 1997. 578-583.
- [2] Booth D, Haas H, McCabe F, Newcomer E, Champion M, Ferris C. Orchard. Web Services Architecture. W3C WG Note[OL]. <http://www.w3.org/TR/ws-arch/>.
- [3] Overhage S. On Specifying Web Services Using UDDI Improvements[A]. 3rd Annual International Conference on Object-Oriented and Internet-based Technologies, Concepts, and Applications for a Networked World Net[C]. ObjectDays: Germany, 2002. 50-61.
- [4] Atkinson C, Stoll D, Acker D. Separating Per-client and Pan-client Views in Service Specification[A]. Proc. Of the Int Workshop on Service Oriented Software Engineering(IW-SOSE)[C]. Shanghai: China, 2006: 80-92.
- [5] Faichamps D. Organizational Factors and Reuse[J]. IEEE Software, 1994, 11(5), 25-36.
- [6] Frakes W B, Fox C J. Quality improvement using a software reuse failure modes model[J]. IEEE Transactions on Software Engineering, 1996, 22(4): 112-120.

- [7] Morisio M, Ezran M, Tully C. Success and failure factors in software reuse[J]. IEEE Transactions on Software Engineering, 2002, 28(4): 340-357
- [8] Rompothong P, Senivongse T. A Query Federation of UDDI Registries[A]. Proceedings of the 1st international symposium on Information and communication technologies[C]. Dublin: Ireland, 2003. 561-566.
- [9] UDDI Specification v3.02. Universal Description Discovery and Integration[OL]. [2007-4-8] <http://www.oasisopen.org/committees/uddi-spec/doc/spec/v3/uddi-v3.0.2-20041019.htm>.
- [10] Tewai V, Dadgee N, Singh I. An Improved Discovery Engine for Efficient and Intelligent Discovery of Web Service with Publication Facility[A]. Services- II [C]. Indore: India, 2009. 63-70.
- [11] Zhenqi Wang, Yuan Yuan Hu. An Approach for Semantic Web Service Discovery Based on P2P Network. [A] Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08. 4th International Conference on [C]. Inf. & Network Manage. Center, North China Electr. Power Univ., Baoding: China. 2008. 1-4.
- [12] Kashani F B, Chen C C and Hahabi C. WSPS: Web Service Peer-to-Peer Discovery Service[A]. In Proceedings of 4th International Conference on Internet Computing[C]. 2004. 733-743
- [13] Banaei-Kashani F, Shahabi C. Searchable Queral Data Networks[A] in Proceedings of the International Workshop on Databases, Information Systems and Peer-to-Peer Computing in conjunction with VLDB'03[C]. Springer: Berlin, 2004: 17-32
- [14] Vanderlei T, Durão F, Martins A. A Cooperative Classification Mechanism for Search and Retrieval of Software Components[A]. Proceedings of the ACM Symposium on Applied Computing (SAC), Information Retrieval Track[C]. ACM New York: USA, 2007. 866-871
- [15] Chunkmol U, Benharkat A N, Amghar Y. Enhancing Web Service Discovery by using Collaborative Tagging System[J]. International Journal of Web Science Practice, 2008, 3(4): 129-135.
- [16] Gupta M, Li R, Y Z. Survey on social tagging techniques[J]. ACM SIGKDD Exploration Newsletter, 2010, 12(1): 650-661
- [17] Elgazzar k, Hassan A E, Martin P. Clustering WSDL Documents to Bootstrap the Discovery of Web Services[A]. ICWS'2010[C]. Miami: Florida, 2010. 147-154.
- [18] Wei Liu, Wilson Wong. Web service clustering using text mining techniques[J]. International Journal of Agent-Oriented Software Engineering, 2009, 3(1): 6-26.
- [19] Atkinson C, Bostan P, Hummel O. A Practical Approach to Web Service Discovery and Retrieval[A]. ICWS[C]. Salt Lake City: Utah, USA, 2007. 241-248.
- [20] Qi Yu and Rege M. On Service Community learning: A Co-clustering Approach[A]. ICWS 2010[C]. Salt Lake City, Utah: USA, 2010. 283-290.
- [21] Dhillon I S. Co-clustering documents and words using bipartite spectral graph partitioning[A]. In KDD '01: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining[C]. New York: NY, USA, 2001. 269-274
- [22] Shi J, Malik J. Normalized cuts and image segmentation[J]. IEEE Trans. Pattern Anal. Mach. Intell., 2000, 22(8): 888-905.
- [23] Chenting Zhao, Ma Chun'e, Jing Zhang. HyperService: Linking and Exploring Services on the Web[A]. ICWS 2010[C]. Salt Lake City: Utah, USA, 2010. 17-24.
- [24] Ricardo A, Baeza Y, Berthier R. Modern Information Retrieval[M]. Addison-Wesley, 1999v
- [25] Salton G. Automatic Text Processing[M]. Addison-Wesley, 1989
- [26] Bultan T, Fu X, Hull R. Conversation Specification: A New Approach to Design and Analysis of E-Service Composition[A]. WWW '03 Proceedings of the 12th international conference on World Wide Web[C]. New York: USA, 2003. 403-410
- [27] Fu X, Bultan T, Su J. Conversation Protocols: A Formalism for Specification and Verification of Reactive Electronic Services[J]. Theoretical Computer Science, 2003, 328(1-2): 19-37.
- [28] BPEL4WS Version 1.1. [2003]. <http://www-106.ibm.com/developerworks/library/ws-bpel/>.
- [29] W3C. WSCL v1.0. [2002]. <http://www.w3.org/TR/2002/NOTE-wscl10-20020314/>.
- [30] Berardi D, Calvanese D, De Giacomo G. Automatic Composition of e-Services that Export their Behavior[A]. Service-Oriented Computing - ICSOC 2003, Trento: Italy, 2003. 43-58
- [31] Conrad Bock. Process Specification Language[OL]. [2003] <http://www.mel.nist.gov/psl/ontology.html>.
- [32] Davulcu H, Kifer M, Ramakrishnan I V. CTR-S: A Logic for Specifying Contracts in Semantic Web Services[A]. WWW Alt. '04 Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters[C]. New York: USA, 2004. 144-153.
- [33] The OWL Coalition. OWL-S: Semantic Markup for Web Services. [2004]. <http://www.daml.org/services/owl-s/1.1B/owl-s.pdf>.
- [34] Shen Z, Su J. Web Service Discovery Based on Behavior Signatures[A]. IEEE International Conference on Services Computing[C]. Dept. of Comput. Sci., California Univ., Santa Barbara, CA: USA, 2005. 279-286.
- [35] Dong X, Halevy A, Madhavan J. Similarity search for web services[A]. VLDB '04 Proceedings of the Thirtieth international conference on Very large data bases - Volume 30[C]. VLDB: Endowment, 2004. 372-383.
- [36] Liu X, Huang G, Mei H. Discovering homogeneous web service community in the user-centric web environment[J]. IEEE T. Services Computing, 2009, 2(2): 167-181.
- [37] Cilibrasi, Rudi L, Vitnyi, Paul M B. The Google similarity distance[J]. IEEE Transactions on Knowledge and Data Engineering, 2007, 19(3): 370-383.
- [38] Sivashanmugam K, Verma K, Sheth A. Adding Semantics to Web Services Standards[A]. in Proceedings of the International Conference on Web Services[C]. New York: USA, 2003. 315-330.
- [39] Xiangfeng Luo, Zheng Xu, Jie Yu and Xue Chen. Building Association Link Network for Semantic Web Resources[J]. 2011, 8(3): 482-494.

- [40] Page L, Brin S, Motwani S. The PageRank citation ranking: Bring order to the web[R]. Technical report: Stanford University, 1998.
- [41] Kleinberg J. Authoritative Sources in a Hyperlinked Environment[J]. Journal of the ACM, 1999, 46(5): 604-622.
- [42] M.F. Porter. An Algorithm for Suffix Stripping[J]. Program, 1980, 14(3): 130-137.