

文章编号:1001-9081(2010)02-0390-04

支持情景应用构建的 Web 服务适配方法

李 婧, 陈旺虎

(西北师范大学 数学与信息科学学院, 兰州 730070)

(nwnlijing@nwnu.edu.cn)

摘 要:为应对情景应用构建过程中,待连接服务之间的不匹配现象,提出了一种半自动的服务适配方法。提出了以树型文法表示服务的消息,并基于树型自动机的代数操作定义了服务的适配操作集,分析了该操作集的完备性。在此基础上,提出了支持情景应用构建的服务适配方法。通过在生物信息情景应用构建场景中的应用,表明该方法具有即时性好、适应服务差异以及无需硬编码的优点。

关键词:情景应用;服务适配;树自动机;树型文法;Web 服务

中图分类号: TP311.52 **文献标志码:** A

Approach to service adaptation for constructing situational applications

LI Jing, CHEN Wang-hu

(College of Mathematics and Information Science, Northwest Normal University, Lanzhou Gansu 730070, China)

Abstract: To solve the mismatch of connected services in situational applications, a semi-automatic service adaptation approach was proposed. Service messages were described by tree grammar and operations for service adaptation were defined based on tree automata algebra. The completion of the set of service adaptation operations was also analyzed. And then, the approach to service adaptation for constructing situational applications was proposed and applied to building situational applications with a scenario of bioinformatics experiments. The results show that the adaptation based on the proposed approach is just-in-time, adaptive to the heterogeneity of services and of no need for hard-coding.

Key words: situational application; service adaptation; tree automata; tree grammar; Web service

0 引言

近年来,应用软件开发呈现出明显的“长尾效应”^[1]。据统计,已有超过100万的Ad Hoc IT从业人员及业务专家试图通过软件应用程序来解决日常工作中碰到的问题^[1]。该类应用通常面向小范围内的一组具有特定需求的用户,具有生命周期短、规模小、非正式以及用户构建等特点^[2],属于典型的情景应用(Situational Applications, SA)。

Web 2.0 以及 SOA 技术的出现进一步推动了情景应用构建方法的发展。用户可通过组装共享的独立组件,获取集成的情景体验。在这些组件中,最为典型的是 Web 服务和 REST(Representational State Transfer)服务^[3],以及 RSS Feed 等,而组装的过程通常基于可视化的环境(如 Mashup Maker^[4])来完成。

在生物信息等科学计算领域,用户需求的多变性和求解问题的探索性^[5],使得情景应用的构建需求日益增加。同时,该领域用于资源共享的 Web 服务的数量目前已超过3 000个^[6],使得基于 SOA 构建该领域的情景应用变得可行。

在基于 SOA 构建情景应用的过程中,如何连接两个 Web 服务是必须面对的核心问题之一,而具有相同业务功能的 Web 服务之间的差异是阻碍该过程透明完成的主要因素。因此,需要解决待连接的 Web 服务之间的不匹配现象,该过程被称为服务适配。

服务适配方面的研究近年来受到了一定的关注。文献[7-10]针对 Web 服务所包含的操作在交互流程上与业务协

议的不匹配现象,提出了一种服务行为的适配方法。由于情景应用往往具有 Informal 的特色^[2],因此,待连接 Web 服务之间的接口适配更受关注。

在 Web 服务接口的适配方面,文献[7,11]将服务在交互过程中的接口不匹配现象分为 Merge/Split 和 Extra/Missing,并基于消息模式(Schema)的映射提出了相应的解决办法。然而,在情景应用的构建过程中,待连接的 Web 服务之间的不匹配现象还包括服务之间传递的消息内容(可简称为消息)。另外,服务消息 Schema 级别的适配往往在应用的构建阶段,需要一定的辅助编码,而情景需求的变化性对适配方法的即时性提出了新的要求。文献[12-13]则基于 Web 的 SOAP 消息的 DTD,通过文法转换规则 XTG,实现服务消息的转换。然而,该方法主要从消息文档的查询角度来完成转换,没有涉及多个服务之间的适配操作,没有关注具有相同业务功能的 Web 服务在描述上的差异等特点。文献[14]提出了一种服务转换代数,从抽象的角度提出了服务在接口转换上的理论基础。

综上所述,本文将研究一种即时性好、无硬编码以及适应 Web 服务描述差异的 Web 服务适配方法,以提高情景应用构建中具有差异的 Web 服务的使用效率以及情景应用的构建效率。

1 Web 服务适配

从本文的研究目标出发,可将 Web 服务的适配定义如下。

定义1 Web 服务适配。指给定待连接的目标 Web 服务

收稿日期:2009-08-06。 基金项目:国家973计划项目(2007CB310805)。

作者简介:李婧(1975-),女,甘肃兰州人,讲师,硕士,主要研究方向:软件工程、服务工程; 陈旺虎(1973-),男,甘肃静宁人,副教授,博士,主要研究方向:服务计算、软件集成。

t 以及源 Web 服务的集合 S , 通过一定的适配手段, 使得 S 的输出消息满足 t 的输入消息的模式 (Schema) 约束, 以及消息内容约束的过程。

文献[11]将服务之间的不匹配现象归结为 Merge/Split 和 Extra/Missing 几种情形。该分类着眼于应用构建时待连接的服务之间在消息 Schema 上的不匹配。然而, 在情景应用的构建过程中, 服务之间在消息的内容传递方面同样会存在不匹配的现象。另一方面, 情景应用构建的即时性、短生命周期以及 Informal 等特点对服务适配方法同样提出了新的要求。

根据我们对生物信息领域内完成序列相似性查找和序列比对的 200 多个 Web 服务的调查, 具有相同业务功能的 Web 服务之间存在很大的差异。从消息模式的角度来看, 具有相同业务功能的 Web 服务的参数个数、参数的取值范围、消息模式的结构等可能不同。其次, Web 服务的隐含状态、参数的取值范围等使得服务在相同的输入情况下, 输出消息的内容也可能存在差异。

因此, Web 服务的适配方法要有效支持情景应用的构建, 需要具有如下特点:

- 1) 支持应用在运行时的适配, 能够处理 Web 服务在消息内容级别上的不匹配现象;
- 2) 适应情景应用构建的即时性以及情景需求的多变性, 适应 Web 服务在描述上的差异, 不能采用硬编码的适配策略;
- 3) 情景应用的 Informal 特性, 允许在服务适配的过程中引入人工的辅助手段。

综上所述, 本文提出在消息内容级别完成服务的适配, 以使该过程具有即时、支持运行时, 以及兼顾消息 Schema 和消息内容的特点。要完成该目标首先要对 Web 服务的消息进行统一描述, 并提供服务消息的基本操纵算子。

2 服务消息的描述及操纵

尽管具有相同业务功能的 Web 服务在具体的描述上存在很大的差异, 但其消息内容的传递均采用 SOAP 协议。SOAP 消息的本质是一个 XML 文档。根据文献[15]的分析, XML 的 Schema 可以采用树自动机进行刻画, 而树自动机的完备的代数操作可以完成 XML 文档的 Schema 以及内容的变换。事实上, 该思想同样适用于 REST 服务, 但本文仅关心 Web 服务。

借鉴文献[15]的思想及其既有成果, 可以给出 Web 服务的消息对应的文档树及其 Schema 的严格定义。

2.1 服务消息的描述

定义 2 Web 服务消息的文档树。令 Σ 和 D 分别表示服务消息文档中的元素类型和基本数据类型的有限集合, 同时空树被标记为 ε ($\varepsilon \in D$), 则一个 Σ 和 D 上的文档树可定义如下。

- 1) ε 是文档树。
- 2) $a(w)$ 是文档树, 其中 $a \in \Sigma, w \in D$ 。
- 3) $a(t_1, t_2, \dots, t_k), k \geq 1$, 是文档树, 其中, $a \in \Sigma, t_i (i = 1, 2, \dots, k)$, 是文档树。
- 4) 没有别的文档树存在。

我们约定, D 中所包含的数据类型为 XML Schema 中内嵌的基本数据类型, 如 XSD: int, XSD: Boolean, XSD: string, XSD: decimal 等。

相应地, Web 服务消息的 Schema 可被定义如下:

定义 3 Web 服务消息的 Schema。服务消息的 Schema 可表示为 $TA = \langle \Sigma, D, Q, \sigma, F \rangle$, 其中:

- 1) Σ, D, Q 分别对应文档的元素类型、数据类型和状态的有限集合;
- 2) σ 是一个函数 $\sigma: \Sigma \times E \rightarrow Q$, 其中, E 可以是 D 的元素 ω , 或者是一个 Q 上的正则表达式 $L(E)$;
- 2) $F \subseteq Q$ 是终止状态集。

由于所有被树自动机 TA 接受的文档树的集合构成了 TA 的语言 $L(M)$, 因此可定义服务的消息如下。

定义 4 给定 Web 服务消息的 Schema M , 所有能被 M 所接受的文档树的集合构成了该 Schema 可接受的消息, 记作 $L(M)$ 。

也就是说, 一个 Web 服务的消息是属于其 Schema 的语言 $L(M)$ 的一棵文档树。

在上述定义的基础上, 可借鉴树自动机的代数操作, 来定义服务消息的基本操纵算子, 并作为实现服务适配操作的元操作。

2.2 服务消息的操纵

将服务的消息看成是一棵文档树, 则其操纵算子依赖于文档树元素构成的正则路径表达式。

定义 5 正则路径表达式。正则路径表达式是字母表 Σ 上的路径的集合, 被定义为 $*.a.*$, 其中 $a \in \Sigma$, $*$ 表示 Σ^* 的任意序列。

路径由从根元素开始的字母表 Σ 的词语的列表构成。在一个文档树中, 一条路径可以表示零个、一个或者多个子树。在一个确定的 TA 中, 一条路径对应一个状态, 而且一个正则路径表达式 P 唯一对应自动机的一个状态子集 Q_P 。

2.2.1 布尔操纵算子

给定两个 Web 服务对应的消息文档树 D_1, D_2 以及文档树上的路径表达式 e , 服务消息的布尔操纵算子并 (Union)、交 (Intersect) 和差 (Difference) 可分别定义如下。

- 1) Union (D_1, D_2, e): 根据路径表达式 e , 获取 D_1 和 D_2 的子树对应的片断, 并对两者进行合并, 结果写入结果文档树。
- 2) Intersect (D_1, D_2, e): 根据路径表达式 e , 获取 D_1 和 D_2 的子树对应的片断, 并求取两者的相同部分, 结果写入结果文档树。
- 3) Difference (D_1, D_2, e): 根据路径表达式 e , 获取 D_1 和 D_2 的子树对应的片断, 并且将 D_1 中不包含在 D_2 中的部分作为结果写入结果文档树。

需要说明的是, 服务消息的文档树的元素可能基于列表 (List-based), 也可能基于集合 (Set-based)。前者将文档树看成是有序树, 而后者将文档树看成是无序树。但根据文献[15]的分析, List-based 的操作可以通过扩展 Set-based 的操作实现。在本文中, 仅讨论 Set-based 的方式。

2.2.2 查询操纵算子

服务消息文档树的查询操纵算子主要包括选择 (Selection)、投影 (Projection Selection)、元素删除 (Element collapsing) 和替换 (Product), 其具体定义如下。

- 1) Select (D, e, c), 其中 D 是一个文档, e 是 Σ 上的一个路径表达式, c 是形如 $e = < value >$ 的条件表达式。Select 操作将 e 指定的 D 的子树看成一个集合, 从中选取满足条件 c 的结果。
- 2) Project (D, e), 其中 D 是一个文档, e 是 Σ 上的一个路径表达式。该操作的运算结果是在 D 中抛弃那些 e 没有涵盖的路径。

与运算的服务消息进行规约,规约的方法见本文 3.3 节。

最后,依赖于服务消息的操作算子,作用于待适配的服务的消息,获取最终的服务适配结果,该结果作为目标服务的输入消息。

4 实验验证和分析

为验证本文方法的有效性,选取了生物信息领域的基因序列组装流程的一个片断进行了实验。由于在基因序列组装的过程中,需要根据当前的实验结果对下一步的行为进行调整,使得实验流程在很多情况下需要进行大量的探索,导致对同类服务的使用在不同情况下有不同的要求,形成了一种典型的情景应用。

在该实验流程中,需要进行序列的相似性查找。试验的上下文要求输入 Fasta 格式的数据,而返回 Person 格式的查找结果。另外,为使序列相似性查找的结果更为准确,要求涉

及 ddbj 和 ipi 等基因数据库。因此,实验过程中需要通过 Fasta 格式转换、序列相似性查找和 Person 格式的格式化三类服务。为方便起见,我们分别标记三类服务在整个应用中所处的位置为节点 A、B 和 C。

试验过程中,共选取了包括用于序列相似性查找、序列格式转换的 Web 服务 100 个。这些 Web 服务来源于 NCBI、EBI 和 DDBJ 等国际权威的生物信息研究机构以及 MyGrid 和 Bimoby 等生物信息研究项目。所有服务部署于上述机构或者项目组的远程 SOAP 服务器上,实验过程中仅读取其描述信息,但不对其作任何修改。另外,基于 MyGrid 项目组的生物信息本体并作适当的修改和扩展,我们对上述服务进行了语义标注。

表 2 给出了在连接 A、B 和 C 三个节点的服务时所采用的服务适配操作的部分结果片段。

表 2 服务适配结果片段

| 节点 A | 节点 B | 节点 C | 适配操作 |
|--------------------------------|--------------------|--------------|------------------------|
| FromFastaToGenericSequence | EBI Blastn | FormatPerson | Merge; Join; Transform |
| FromFastaToGenericSequence | RunNCBIBlastn | FormatPerson | Merge; Transform |
| FromFastaToGenericSequence | Blastlocal-blastn | FormatPerson | Join; Transform |
| FromFastaToGenericSequence | DDBJ Blast | FormatPerson | Extraction; Transform |
| FromFastaToGenericSequence | Timelogical-blastn | FormatPerson | Join; Transform |
| FromFastaToCommentedNASequence | EBI Blastn | FormatPerson | Merge; Join; Transform |
| FromFastaToNucleotideSequence | EBI Blastn | FormatPerson | Merge; Join; Transform |
| ... | ... | ... | ... |

从表 2 可看出:节点 A 选择 FromFastaToGenericSequence 服务的时候,节点 B 可选择 EBI、DDBJ 等多个 Blastn 服务以及 Blast 服务。这些服务的参数个数存在很大差异;同时,一些服务使用简单数据类型,而一些服务使用复杂数据类型,导致了这些服务在 Schema 上存在很大差异。另外,如果指定核苷酸查找,则 Blast 服务的部分结果不满足要求,而如果指定同时包含 ddbj 和 ipi 序列数据库,则单个的 EBI 或者 NCBI 的 Blastn 服务均不满足要求。也就是说,A 和 B 在消息内容上同样存在不匹配。从表 2 中可以看出,由于本文的服务适配方法基于服务的消息来完成,因此兼顾了 Schema 级别和消息内容级别的适配,并在使用所定义的服务适配操作的基础上,完成了整个服务适配过程。

图 2 分析了在实验过程中,对本文所提出的适配操作的总体应用状况,以及需要人工辅助来完成的适配操作的比例。

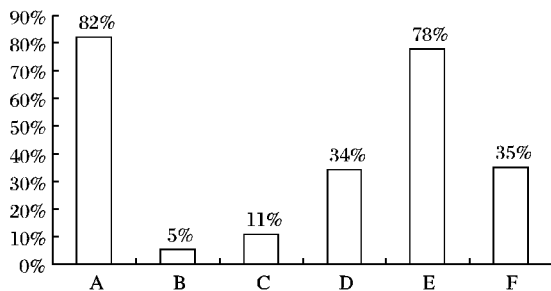


图 2 适配操作的应用状况分析

在图 2 中,A 为总适配率,表示被适配服务和总服务数的比值;B、C、D 和 E 分别表示了 Merge、Extract、Join 和 Transform 在所有适配操作中所占的比例;E 为人工辅助率,表示需要人工参与的适配操作数和总适配操作数的比值。从图 2 可以看出,在情景应用的构建过程中,大约 82% 的服务连接过程需

要通过适配来完成,其中,需要对消息内容进行合并和抽取转换的分别占到了 5% 和 11%,说明了本文方法基于消息完成适配的重要性。另外,可以看出需要手工辅助的适配操作约占 35%,从而说明本文方法的半自动性,可在一定程度上提高服务适配的效率。

从实验结果中可以看出,本文的服务适配方法避免了基于服务返回对象的变换过程中变换操作的不兼容,从而可适应 Web 服务的接口差异性。同时,本文方法使得服务适配过程具有很好的即时性,并且依赖基于服务消息的适配操作,使得适配过程无需在应用构建阶段进行硬编码。

5 结语

本文提出了一种半自动的支持情景应用构建的服务适配方法,该方法具有即时性好、无硬编码以及适应 Web 服务差异的特点。实验和分析表明了该方法的可行性和有效性。

参考文献:

- [1] JHINGRAN A, JOSE S. Enterprise information mashups: Integrating information, simply[C]// Proceedings of the 32nd International Conference on Very Large Data Bases. New York: ACM, 2006: 3-4.
- [2] De WEVER B, MECHANT P, VEEVAETE P, et al. E-learning 2.0: Social software for educational use[C]//Proceedings of the 9th IEEE International Symposium on Multimedia. Washington, DC: IEEE Computer Society, 2007: 511-516.
- [3] PAUTASSO C. RESTful Web service composition with BPEL for REST[J]. Data & Knowledge Engineering, 2009, 68(9): 851-866.
- [4] YU J, BENATALLAH B, CASATI F, et al. Understanding mashup development[J]. IEEE Internet Computing, 2008, 12(5): 44-52.

(下转第 410 页)

解决了航天器数据查询与分析系统灵活性和扩展性差的问题。该方法已经在航天器试验与测试数据查询分析平台中得到了应用,并可方便地移植到其他企业应用系统中。

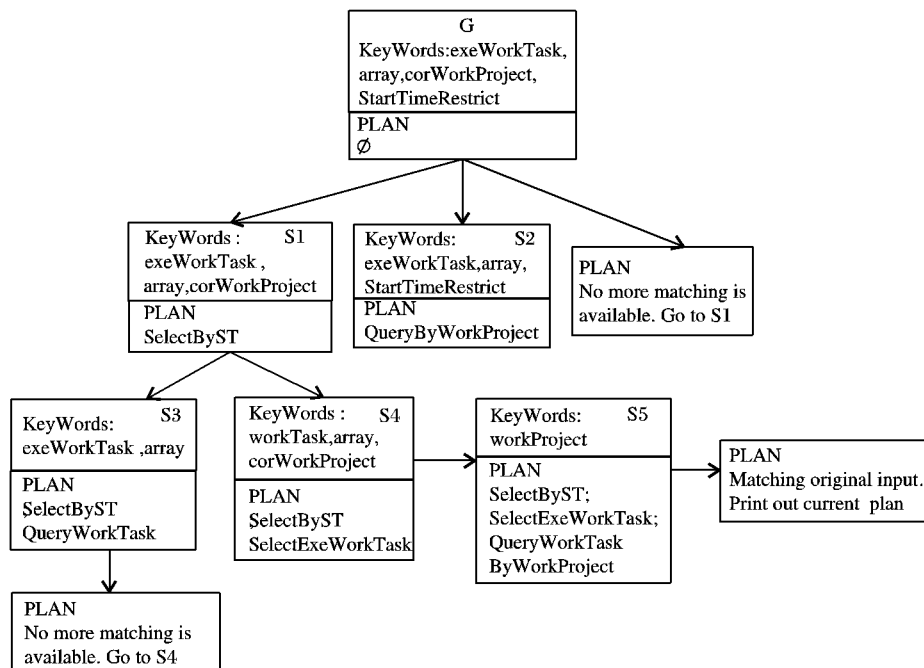


图6 后向搜索算法工作过程

参考文献:

- [1] 万常选, 鲁远. 基于权重查询词的 XML 结构查询扩展[J]. 软件学报, 2008, 19(10): 2611-2619.
- [2] RIABOV A V, R BOUILLET E, FEBLOWITZ M B, *et al.* Wishful search: Interactive composition of data mashups[C]// 17th International World Wide Web Conference. New York: ACM, 2008: 775-784.
- [3] BOUILLET E, FEBLOWITZ M, ZHEN L. *et al.* A tag-based approach for the design and composition of information processing applications[C]// Object Oriented Programming, Systems, Languages and Applications. New York: ACM, 2008: 585-602.
- [4] ZHEN L, RANGANATHAN A, RIABOV A, *et al.* A planning-based approach for the automated configuration of the enterprise service bus [C]// International Conference on Service-Oriented Computing. Berlin: Springer, 2008: 538-544.
- [5] BOUILLET E, FEBLOWITZ M, FENG H, *et al.* A folksonomy-based model of Web services for discovery and automatic composition [C]// Proceedings of the 2008 IEEE International Conference on Services Computing. Washington, DC: IEEE, 2008, 1: 389-396.
- [6] BIZER C, CYGANIAK R, GAUSS T. The RDF book mashup: From Web APIs to a Web of data[EB/OL]. [2009-06-01]. <http://www.dfki.uni-kl.de/~grimnes/2007/06/SFSW07Papers/6.pdf>
- [7] CHARIF Y, SABOURET N. An overview of semantic Web services composition approaches [EB/OL]. [2009-08-01]. <http://www.poleia.lip6.fr/~sabouret/ps/Charif-Sabouret-Context2005.pdf>.
- [8] YU D, LI Z W, YE G, *et al.* The workflow-based modeling method for spacecraft automatic testing process [C]// International Conference on Information and Knowledge Engineering. Las Vegas: CSREA Press, 2008: 267-271.
- [9] RIABOV A, LIU Z. Planning for stream processing systems[C]// Proceedings of the 20th National Conference on Artificial Intelligence. Pittsburgh: AAAI Press, 2005, 1: 1205-1210.
- [10] GHALLAB M, HOWE A, KNOBLOCK C, *et al.* PDDL: The planning domain definition language. [EB/OL]. [2008-10-10]. <http://www.cs.yale.edu/pub/mcdermott/software/pddl.tar.gz>.
- [11] BYLANDER T. The computational complexity of propositional STRIPS planning[J]. Artificial Intelligence, 1994, 69(1/2): 165-204.
- [5] 陈旺虎, 李婧. 基于抽象服务模型提高 Web 服务匹配的适应性[J]. 计算机科学, 2009, 36(9): 139-142.
- [6] HULL D, ZOLIN E, BOYVKIN A, *et al.* Deciding semantic matching of stateless services[C]// Proceedings of the Twenty-First National Conference on Artificial Intelligence. California: AAAI Press, 2006: 16-20.
- [7] MOTAHARI H R, BENATALLAH B, MARTENS A, *et al.* Semi-automated adaptation of service interactions [C]// WWW'07: Proceedings of the 16th International Conference on World Wide Web. New York: ACM Press, 2007: 993-1002.
- [8] BROGI R, POPESCU. Automated generation of BPEL adapters [C]// Proceedings of the 4th International Conference on Services Oriented Computing. New York: Springer-Verlag, 2006: 27-39.
- [9] KONGDENFHA W, SAINT-PAUL R, BENATALLAH B, *et al.* An aspect-oriented framework for service adaptation[C]// Proceedings of the 4th International Conference on Service Oriented Computing, Chicago, USA, 2006. New York: Springer-Verlag, 2006: 15-26.
- [10] HEMER D. A formal approach to component adaptation and composition[C]// Proceedings of the 28th Australasian Conference on Computer Science. Darlinghurst: Australian Computer Society, 2005: 259-266.
- [11] DUMAS M, SPORK M, WANG K. Adapt or perish: Algebra and visual notation for service interface adaptation[C]// Proceedings of the 4th International Conference on Business Process Management. Berlin: Springer, 2006: 65-80.
- [12] WANG X, CHEUNG S. C. Grammar based interface processing in Web service composition [C]// Proceedings of Conceptual Modeling for Advanced Application Domains. New York: Springer-Verlag, 2004: 323-334.
- [13] 王晓玲, 郭志懋, 周傲英. Web 服务组合的基于文法的消息处理. 计算机学报, 2005, 28(4): 478-485.
- [14] 房俊. 支持最终用户编程的服务虚拟化方法研究[D]. 北京: 中国科学院计算技术研究所, 2006.
- [15] CHIDLOVSKII B. Using regular tree automata as XML schemas [C]// Proceedings of IEEE Advances in Digital Libraries 2000. Washington, DC: IEEE Computer Society, 2000: 89-104.
- [16] 陈旺虎, 刘晨, 李厚福, 等. 支持虚拟组织的语义基础设施的动态构建方法研究[J]. 计算机学报, 2006, 29(7): 1127-1136.
- [17] 喻坚, 韩燕波. 面向服务的计算——原理和应用[M]. 北京: 清华大学出版社, 2006: 44-53.

(上接第393页)