

基于 SOAP 报文组装的生物信息学 Web 服务客户端

闵波 杨帆 颜炜 杜生 雷权 施雪伟

(新疆军区总医院 乌鲁木齐 830000)

〔摘要〕 介绍 Web 服务及 SOAP 报文动态组装的内涵与过程, 基于 SOAP 报文组装的方法构建生物信息学 Web 服务客户端, 介绍系统设计、实现与应用测试, 提供 Web 服务动态解析、调用及数据管理的可视化操作接口, 提升用户利用 Web 服务资源的效率。

〔关键词〕 生物信息学; Web 服务动态调用; SOAP 报文组装

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2018.07.008

Bioinformatics Web Service Client Assembled Based on SOAP Message MIN Bo, YANG Fang, YAN Wei, DU Sheng, LEI Quan, SHI Xue-wei, General Hospital of Xinjiang Military Region, Urumqi 830000, China

〔Abstract〕 The paper introduces the connotation and procedure of Web service and the dynamic assembly of SOAP message. The bioinformatics Web service client is built based on assembly method of SOAP message, it introduces system design, realization and application test, provides Web service dynamic analysis, invoking and the visualized operation interface of data management, as well as enhances users' efficiency in using Web service resources.

〔Keywords〕 Bioinformatics; Web service invoke dynamically; SOAP message assembly

1 引言

随着互联网的高速发展, 大量生物信息数据库服务提供 Web 服务接口, 帮助用户通过程序调用方式快速访问数据和工具资源。Web 服务一方面克服人工检索数据库的繁琐, 提高资源访问和获取效率, 另一方面提供一种解决数据异构与整合问题的可能途径^[1]。为有效地利用和管理生物信息学社区丰富的 Web 服务资源, 构建 Web 服务客户端具有

重要意义。

由于生物数据的多样性与复杂性, 生物信息数据库通常是建立在独立的学术或商业机构, 这些不同的服务提供者定义的 Web 服务普遍存在描述结构、语法和语义异构等问题。因此许多生物信息学家尝试构建 Web 服务客户端, 其具备良好的兼容性与扩展性, 以便整合尽可能多的 Web 服务资源。近年来生物信息学社区已经出现大量 Web 服务客户端工具, 如 INB—Client^[2] 和 jORCA^[3] 等, 通过对多种数据类型的兼容支持, 能够动态地访问多种类型的 Web 服务, 但该方法容易导致用户提供的参数难以动态匹配服务定义参数类型, 使客户端兼容的服务资源有限, 通用性和扩展性不足。为克服动态绑定的不足, 本研究基于简单对象访问协议 (Simple Object

〔收稿日期〕 2018-04-02

〔作者简介〕 闵波, 技师, 发表论文 11 篇; 通讯作者: 杨帆, 主治医师。

Access Protocol, SOAP) 报文组装的方法构建生物信息学 Web 服务客户端平台, 帮助用户实现对 Web 服务动态解析、调用及数据的统一可视化管理。

2 Web 服务

2.1 概述

Web 服务是一个平台独立、低耦合、自包含的 Web 应用程序编程接口 (Application Programming Interface, API), 服务请求者可以通过构建客户端程序来调用服务提供者开放的 Web 服务。Web 服务使用可扩展标记语言 (Extensible Markup Language, XML) 与超文本传输协议 (Hyper Text Transfer Protocol, HTTP) 的 SOAP 协议将信息编码成 XML 格式的 SOAP 报文进行数据交换, 使不同技术实现的应用程序在不同操作系统间能够互相通信^[4]。为用户公开 Web 服务功能与函数调用的参数信息, 服务提供者利用 Web 服务描述语言 (Web Service Description Language, WSDL) 文档将服务的部署实现信息抽象定义为一组服务访问点。

2.2 SOAP 报文与 WSDL 文档结构

2.2.1 SOAP 报文 SOAP 是一种基于 XML 的简易对象访问协议, 可使应用程序在 HTTP 上进行信息交换, SOAP 报文实际上是描述传递信息的 XML 文档, 主要包括 Envelope、Header、Body 及 Fault 元素, 其中 Envelope 与 Body 元素是报文必需的核心组件, Body 元素包含需传送到服务终端的实体消息, Envelope 元素用于将 XML 文档标识为 SOAP 消息。Header 元素与 Fault 元素是可选的, Header 元素主要包含报文的头部信息, Fault 元素提供处理消息时发生错误的信息。

2.2.2 WSDL 文档 与 SOAP 报文类似, WSDL 文档是一种用于描述 Web 服务以及如何访问该服务的 XML 文档, 主要包括 Types、PortType、Message 和 Binding 4 个必需元素以及可选的 Service 元素。PortType 元素描述一个或多个端点支持的操作抽象集, Binding 元素定义特定端口类型的具体协议和数据格

式规范, Message 元素定义要传递的数据, Types 元素定义服务使用的数据类型, Service 元素主要定义相关服务访问点的集合。

2.3 Web 服务客户端

主要用于实现对 Web 服务接口的远程调用, 基本框架, 见图 1。客户端首先对服务提供者发布的 WSDL 文档进行解析, 获取描述 Web 服务的函数、参数和返回值等信息, 进而调用相应的服务程序接口。客户端调用 Web 服务包括静态调用和动态调用两种方法, 静态调用主要指在用户程序中设置已知的服务访问信息来远程调用, 或者利用工具将 WSDL 文档序列化成客户端接口类, 然后直接对接口类进行本地调用^[5]。静态调用过多地涉及人为参与, 效率较低, 因此大多数开发者习惯使用动态调用的方法。动态调用是指在用户完全不了解 Web 服务组成时, 动态地发现或调用所需的 Web 服务, 在服务消费者与提供者之间进行解耦^[6]。然而客户端在 Web 服务动态调用过程中, 由于 WSDL 文档存在自定义的数据结构类型, 客户端难以直接将用户提交的数据动态绑定为服务定义的参数类型, 导致大多数客户端兼容的 Web 服务资源有限。利用 XML 序列化工具根据 WSDL 文档动态生成相应静态类的方法, 虽然能够解决参数类型动态绑定的问题, 但客户端整合的服务越多序列化生成的类的数量也越多, 进而导致程序代码臃肿, 不便管理, 增加客户端的复杂性。为能够解决这种兼容性与复杂性问题, 本研究提出基于 SOAP 报文组装来构建客户端的方法。

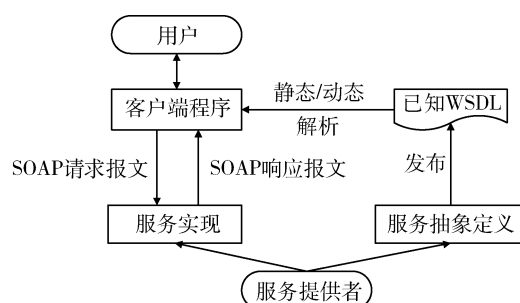


图 1 Web 服务客户端调用基本框架

3 SOAP 报文动态组装

3.1 概述

指直接解析 WSDL 文档定义的服务信息, 根据特定的函数调用, 动态生成相应的 SOAP 报文。SOAP 报文组装具体包括 WSDL 文档解析、XML Schema 数据类型定义的转化以及组装报文中 Body 元素的 XML 文档内容 3 大组件。常规 XML 解析器可以方便地解析 WSDL 文档, 获取服务定义的各种信息, 因此 SOAP 报文动态组装的重点是如何完成复杂类型定义的转化及 Body 元素的组装。

3.2 复杂类型定义转化成 XML 数据

XML Schema 中定义的复杂类型主要指包含子元素、空元素或混合内容的元素。通过解析 WSDL 文档的数据类型定义 Schema, 将其中的复杂数据类型转化成空的 XML 元素对象。整个转化过程为: 首先遍历所有复杂数据类型节点; 其次解析节点子元素的顺序、名称与类型, 如果子元素为叶子节点, 或子元素的类型定义为引用其他复杂类型, 则进一步对子元素进行递归解析; 最后序列化复杂数据类型, 生成对应的 XML 数据文档。

3.3 Body 元素的 XML 文档内容组装

选取某个特定函数, 然后建立空的 SOAP 报文对象, 在 Body 元素后添加函数信息元素, 在函数元素后添加参数元素。WSDL 文档包括 4 种不同样式, 分别为文档/文字、文档/编码、RPC/文字及 RPC/编码, 由于不同样式定义了不同的 SOAP 报文编码格式, 需组装成不同格式的 SOAP 报文。如果函数的参数类型是 Schema 中定义的复杂数据类型, 则将参数类型元素替换为相应的 XML 元素对象。组装好的抽象 SOAP 报文进一步接收用户的输入, 向 Body 元素的函数参数传入参数值, 生成包含用户数据的 SOAP 请求报文, 然后发送 SOAP 请求至

服务提供者, 完成对服务的调用。SOAP 报文动态组装的方法对用户来说, 提交的数据不必动态匹配 WSDL 描述定义的参数类型, 避免动态创建参数类型对象的过程, 克服传统服务调用中参数难以动态绑定的瓶颈。

4 生物信息学 Web 服务客户端构建

4.1 概述

生物信息学社区各类 Web 服务资源数量多且分布广, 可基于 SOAP 报文动态组装的方法构建 Web 服务客户端, 提供统一的图形交互接口, 帮助生物信息学家实现对 Web 服务资源的高效管理与使用。

4.2 系统设计

客户端总体设计, 见图 2。主要包括服务检索、WSDL 解析、报文组装和输出重组 4 个组成部分, 通过统一的图形交互接口呈现信息与可视操作。其中 WSDL 解析与报文组装是客户端的核心组件, 两者封装所有服务调用的业务逻辑。服务检索用于方便用户快速查询所需服务。输出重组提供用户便于理解的存储格式和可视化显示, 由于不同操作方法返回的数据格式多种多样, 无法实现对所有种类的数据进行解析组织, 可以调用一些外部可视化程序对数据进行呈现。图形用户接口主要用于实现操作 XML 格式 SOAP 报文的可视交互接口, 方便用户快速地完成报文编辑, 提升客户端的可操作性和友好性。系统利用树形结构来呈现 SOAP 报文, 统一将 XML 文档的元素映射到树中的节点, 将报文的元素绘制成相应的树节点, 再以不同的图形节点方式显示。树的路径对应 XML 文档的 Xpath 路径, 树的操作被动态地序列化成 XML 文档的操作, 通过对树节点的快速遍历与定位, 实现对不同 SOAP 报文元素的编辑赋值。对用户来说整个底层 SOAP 报文的 XML 传值操作是可视化的交互过程。

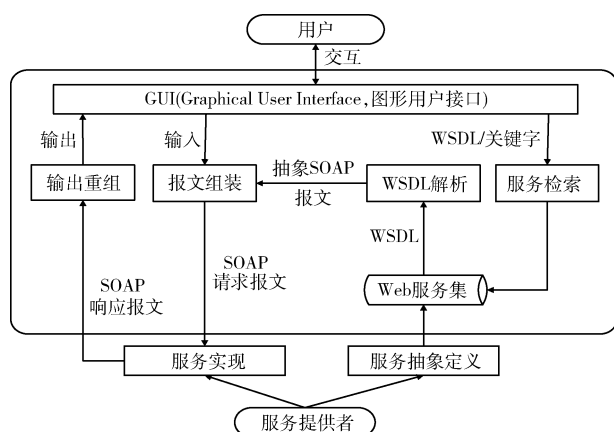


图 2 客户端总体设计

4.3 系统实现与应用测试

4.3.1 系统实现 基于 Java 平台开发, 整合引用包括 WSDL4J、SAA4J、HttpClient、XMLBeans 等核心类库。用户提交指定的 Web 服务 WSDL 文档或统一资源定位地址 (Uniform Resource Locator, URL), 客户端装载并动态解析 WSDL, 不同的操作函数生成相应的抽象 SOAP 报文, 然后用户对抽象报文进行数据填充, 组装为调用服务的 SOAP 请求报文并发送给 Web 服务核心执行组件。服务执行结束后, 客户端接收返回的响应 SOAP 消息并对返回的消息进行解析, 再按特定格式进行重新组织, 最后将结果格式化输出给用户。

4.3.2 应用测试 以京都基因与基因组百科全书 (Kyoto Encyclopedia of Genes and Genomes, KEGG, <http://www.kegg.jp/>) Web 服务为例, 该服务允许用户基于 KEGG 数据库定制生物信息学分析任务, 如检索和计算细胞生化通路。导入 KEGG 的 WSDL 文档, 解析后选取 `get_html_of_colored_pathway_by_objects` 操作函数, 函数的输入参数包含 3 个数组对象, WSDL 文档中定义的报文编码样式为 RPC/Encoded。客户端组装 SOAP 报文实例, 见图 3。通过左边的报文编辑窗口, 用户对生成的抽象报文输入数据, 填充数据后的 SOAP 报文如右边所示, 然后向服务提供者发送该 SOAP 请求报文, 完成服务调用。传统客户端对这类函数传值必须将用户的输入数据动态绑定为数组类型对象, 然后动态地将输入转换成对应 SOAP 请求, 完成该过程必须在调用前静态地声明类型, 否则遇到服务提供者自定义类型难以实现动态调用。如 EBI 中的 InterProScan 服务, 其中 `run` 操作函数的输入参数包括 `email`、`title` 和 `parameters` 3 种, `parameters` 在 WSDL 文档中 Schema 被定义为 `InputParameters` 复杂数据类型, 此时客户端很难自动识别这类自定义的参数类型对象。本客户端利用 SOAP 报文组装的方法生成参数 `parameters` 的报文形式, 见图 4。用户填充报文便可实现对参数 `parameters` 赋值。

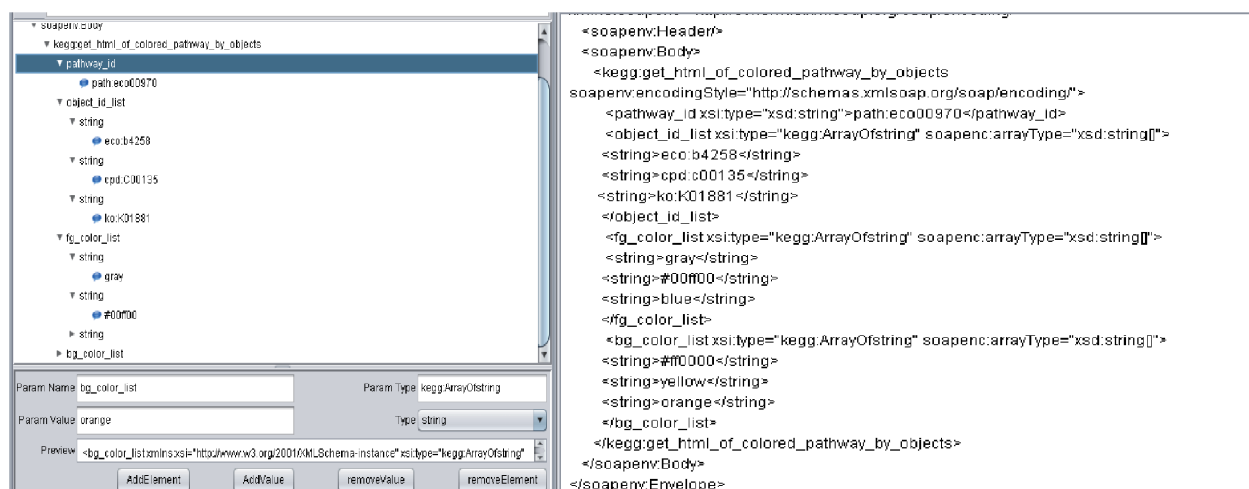


图 3 SOAP 报文组装实例

```

<parameters>
  <nocr>null</nocr>
  <goterms>null</goterms>
  <appl>
    <string>null</string>
  </appl>
  <sequence>null</sequence>
</parameters>

```

图 4 EBI 中 iprscan Web 服务 WSDL 中参数 parameters 的报文

5 结语

当前生物信息学数据资源与分析服务已经成为生命科学研究非常有效的方法与工具,而设计开发新的数据管理与分析工具更是生物信息学家一项重要研究内容。本研究基于 SOAP 报文组装的方法构建生物信息学 Web 服务客户端,相比其他生物信息学社区的 Web 服务客户端,其通用性更好,兼容的 Web 服务资源覆盖面更广。此外客户端提供 SOAP 报文图形化交互接口,这对于部分缺乏编程知识的生命科学研究人员来说具有较好的便捷性、友好性、可操作性。总体来说直接封装用户的输入来创建 SOAP 报文虽然克服了动态创建参数类型对象的瓶颈,但也带来服务动态解析与报文封装操作的复杂性。从理论上说本客户端的方法能够兼容绝大多数 Web 服务资源,但客户端程序能够动态地匹配所

有服务,自动地理解服务提供者和请求者之间传递的消息数据仍只是一种理想的状态^[7]。因此要构建真正通用的生物信息学 Web 服务客户端,需要领域专家共同努力,建立一套标准的机制和数据格式,使不同生物信息学机构实现互相兼容的 Web 服务。

参考文献

- 1 Heinz Stockinger, Teresa Attwood, Shahid Nadeem Chohan, et al. Experience Using Web Services for Biological Sequence Analysis [J]. Briefings in bioinformatics, 2008, 9 (6): 493 - 505.
- 2 Ismael Navas - Delgado, Maria del Mar Rojano - Munoz, Sergio Raml rez, et al. Intelligent Client for Integrating Bioinformatics Services [J]. Bioinformatics, 2006, 22 (1): 106 - 111.
- 3 Victoria Martín - Requena, Javier Ríos, Maximiliano García, et al. JORCA: easily integrating bioinformatics Web Services [J]. Bioinformatics, 2010, 4 (26): 553 - 559.
- 4 周元哲,王荣喜. 基于 SOAP 和 Axis2 的 Web 服务设计与实现 [J]. 西安文理学院学报 (自然科学版), 2010, 13 (2): 64 - 67.
- 5 于建军,孙鹏. 一种 Web Service 动态调用框架的设计与实现 [J]. 计算机工程, 2006, 32 (4): 119 - 121.
- 6 何红波,熊桂林,李义兵,等. 基于 Web Services 的生物信息集成框架 [J]. 系统工程, 2004, 22 (2): 103 - 106.
- 7 李苏,刘建勋. Web 服务的绑定与调用方法研究 [J]. 计算机技术与发展, 2010, 20 (6): 59 - 63.

(上接第 26 页)

- 23 国务院.《新一代人工智能发展规划的通知》[EB/OL]. [2017 - 07 - 20]. http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm.
- 24 中国大数据. 习近平:实施国家大数据战略,加快建设数字中国 [EB/OL]. [2017 - 12 - 12]. <http://www.thebigdata.cn/YeJieDongTai/35064.html>.
- 25 搜狐网. 2017 年末中国大陆总人口超 13.9 亿比上年末增 737 万人 [EB/OL]. [2018 - 01 - 18]. http://www.sohu.com/a/217499248_118392.
- 26 范月丹,潘烁,李俊,等. 医疗大数据应用 SWOT 分析 [J]. 医学信息学杂志, 2016, 37 (7): 51 - 56.
- 27 杜欣,曾伟杰,李承炜,等. 基于移动医疗的孕产妇健康监护系统 [J]. 生物医学工程学杂志, 2016, 33 (1): 2 - 7.
- 28 侯娟茹,周珊,张红梅,等. 达芬奇手术机器人的应用

- 进展与前景展望 [J]. 现代医学与健康研究电子杂志, 2017, 1 (4): 161.
- 29 洪延青,何延哲. 英国健康医疗大数据平台 care. data 为何停摆? [J]. 中国经济周刊, 2016, (29): 77 - 79.
- 30 蔡佳慧,张涛,宗文红. 医疗大数据面临的挑战及思考 [J]. 中国卫生信息管理杂志, 2013, 10 (4): 292 - 295.
- 31 游顶云,蔡乐. 云南省健康医疗大数据的应用前景、挑战及对策研究 [J]. 云南科技管理, 2017, 30 (1): 14 - 16.
- 32 何耀平,叶坚骏,陈刚,等. 地市级医疗健康大数据一级共享平台构建研究与实践 [J]. 中国数字医学, 2016, 11 (8): 34 - 37.
- 33 刘辉,丛亚丽. 临床医学大数据的伦理问题初探 [J]. 医学与哲学 (A), 2016, 37 (10): 32 - 36.