

# 基于 Web Services 的生物信息解决方案<sup>\*</sup>

林毅申, 林丕源

(华南农业大学 计算机科学与工程系, 广东 广州 510642)

摘 要: 在生物信息学系统设计中引入 Web Services 体系结构, 提出生物信息系统功能服务的模型。以双序列比对作为例子, 完成了 Web Services 的设计和实现, 体现出 Web Services 在生物信息学系统应用中的优越性。

关键词: Web Services; 生物信息学; 序列比对

中图法分类号: TP393. 09      文献标识码: A      文章编号: 1001-3695(2005)06-0157-02

## A Solution of Bioinformatical System Based on Web Services

LIN Yi-shen, LIN Pi-yuan

(Dept. of Computer Science & Engineering, South China Agricultural University, Guangzhou Guangdong 510642, China)

**Abstract:** The architecture of Web Services is introduced into the bioinformatics system design, and a bioinformatical service system model is proposed. Using the example of two-sequence alignment, a Web Service is designed and implemented. It shows the advantages of Web Services used in bioinformatics applications.

**Key words:** Web Services; Bioinformatics; Alignment

Web Services 是当今 IT 应用的焦点之一, 它的主要目标是在现有的各种异构平台的基础上构成一个通用的与平台无关、语言无关的技术层, 各种不同平台之上的应用依靠该技术层来实施彼此的连接和集成。传统 Web 应用技术解决的问题是: 如何让人来使用 Web 应用所提供的服务, 而 Web Services 则要解决如何让计算机系统来使用 Web 应用所提供的服务<sup>[1]</sup>。具体到生物信息学系统中, 就是怎样可以封装算法, 简化程序调用过程和提高代码的可重用性。

生物信息学(Bioinformatics)是一门综合计算机科学和生物学等多方面知识的新兴的交叉学科。20 世纪后期分子生物学的长足进展, 把生命活动的物质基础追溯到核酸和蛋白质两大类生物大分子的序列, 它们构成了生物数据的主要部分。对于生物大分子结构、相互作用和生物功能的研究, 也产生着大量数据。正是由于有大量的生物学数据需要搜集和分析, 因此生物信息学是近来比较热门的研究领域<sup>[2]</sup>。生物信息学在计算机领域所涉及的研究主要包括生物信息数据库的建立、数据库接口和检查工具的研制、新的算法和软件研究等三大方面。目前, 我们所做的主要工作是二次数据库的建立以及数据库接口和检查工具的研制, 建设有农业特色的生物信息系统。

### 1 Web Services 体系结构

Web Services 是一种新的 Web 应用程序分支, 它们是自包含、自描述、模块化的应用, 可以发布、定位以及通过 Web 调用<sup>[1,3]</sup>。Web Services 可以执行从简单的请求到复杂商务处理的任何功能。一个 Web Service 一旦部署以后, 其他 Web Services 和应用程序可以发现并调用它部署的服务, 也就是由计算机系统使用 Web 应用所提供的服务。Web Services 的迷人

之处在于: 它可以使用标准的互联网协议, 将功能提纲挈领地体现在互联网和企业内部网上。从程序设计角度看, 可以把 Web Services 视作 Web 上组件的程序设计。

最新公布的 Web Services 体系结构<sup>[3]</sup>如图 1 所示。服务提供者(Provider)和服务请求者(Requester)作为两个独立的实体, 是构成 Web Services 应用的基本元素。Web Services 的工作原理是: 在两个实体间建立连接, 两者就语义(Semantics), Web Services 描述(WSD)达成一致。使用 Web 服务是通过代理(Agent)进行的, 从程序设计角度就是建立代理类, 映射所调用的 Web 服务, 使本地程序可以直接使用。服务请求者与提供者进行交互(Interact)完成数据交换。

Web Services 标准具体到工具和企业标准, 目前分为微软的 .NET 平台和 SUN 公司的 J2EE 标准两种、它们在程序设计环境、语法结构、管理配置等方面也各有自己的特色<sup>[4]</sup>。

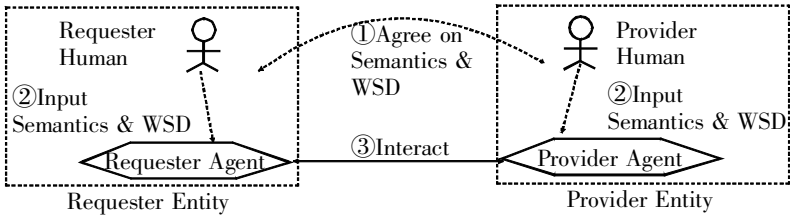


图 1 Web Services 体系结构

### 2 基于 Web Services 的生物信息系统模型设计

#### 2.1 一般的生物信息系统模型

在生物信息研究中, 对基因或蛋白质的研究一般遵循这样的步骤: 在数据库中检索符合条件的数据(基因或蛋白质序列), 接着选择并获取指定的数据, 然后使用特定算法对数据进行处理, 最后根据结果进行人工分析和判断<sup>[2]</sup>。根据数据库和实现算法存放位置的不同, 可以把生物信息系统分为本地系统(Local Bioinformatics System)和远程系统(Remote Bioinfor-

matics System) 两大类。

在本地系统中,数据以文件形式或二次数据库形式存放在本地。在生物信息研究中,往往需要大量的数据素材,本地数据可能不能满足需求,通常要连接到 Internet 获取新数据。另外由于某些算法(如序列比对算法)需要大量计算资源,受限于本地机器的配置,程序运算速度难以忍受。但是,本地系统大多数专长于某个特定领域或特定功能分析,在生物信息研究中是必不可少的。本地生物信息系统模型如图 2 所示。

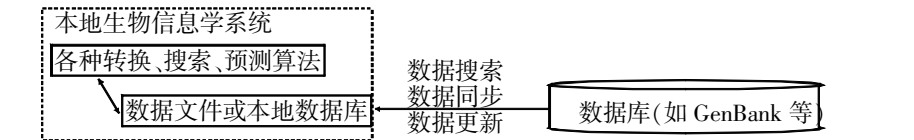


图 2 本地生物信息系统结构

远程系统通常结合专业数据库,有大量专业数据支持。在性能上,大量采用大型机和集群等高性能设备,运算速度极快。但由于远程系统主要针对大众用户提供服务,所提供的功能有限,不能完全满足用户的需要。因此,用户一般使用远程系统进行数据搜索和比对,把得出的结果再放回本地系统作进一步处理。远程生物信息系统模型如图 3 所示。

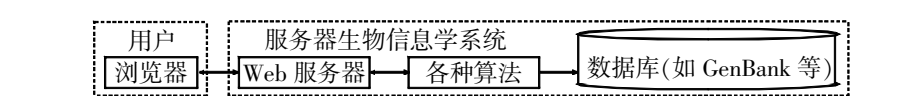


图 3 远程生物信息系统结构

### 2.2 基于 Web Services 的生物信息学系统模型

结合以上系统的本地和远程生物信息系统模型的优点,利用 Web Services 思想对生物信息应用重新设计,我们发现 Web Services 恰好能弥补两个传统模型的缺点。图 4 是我们提出的基于 Web Services 的生物信息系统模型。



图 4 基于 Web Services 的生物信息系统模型

在此模型中,把算法部分封装在 Web Services 中,对外只提供用户调用算法接口和数据库访问接口。用户调用算法接口是为用户提供一个标准数据交换格式,用户按规范就可以使用此 Web Services,不用了解具体的算法实现过程。数据库访问接口是 Web Services 和其他数据库进行数据交换的接口,一般封装在 Web Services 中,对用户而言是透明的。

采用此模型设计,对设计者(一般是计算机学家)而言,有了一个编程的规范,从而使代码重用成为可能;对使用者(一般是生物学家)而言,只要按说明构造一些简单的调用,就可以在高性能机器上进行一些专业的研究。为此,我们把这个基于 Web Services 的生物信息系统称为生物信息系统功能服务模式。此模型既解决了本地系统计算资源的缺乏和获取数据的困难,又解决了远程系统缺乏专业功能的缺点。

## 3 基于 Web Services 的双序列比对方案的实现

生物信息系统中,序列比对是生物信息研究的基础。通过序列比对,可以进行同源核酸或蛋白质序列的查找、引物或探针特异性分析、物种进化树的构建等<sup>[2]</sup>。下面按照生物信息系统功能服务模式的设计对双序列比对功能进行具体实现。

### 3.1 双序列比对流程分析

在双序列比对中,一般的步骤是:由用户指定要进行比对

的序列编号,程序根据此编号从数据库中查找序列内容。获取序列数据后,使用一定的算法(如双序列比对的动态规划算法)对序列进行处理,最后把比对的结果返回给用户。整个流程使用 UML 描述如图 5 所示。

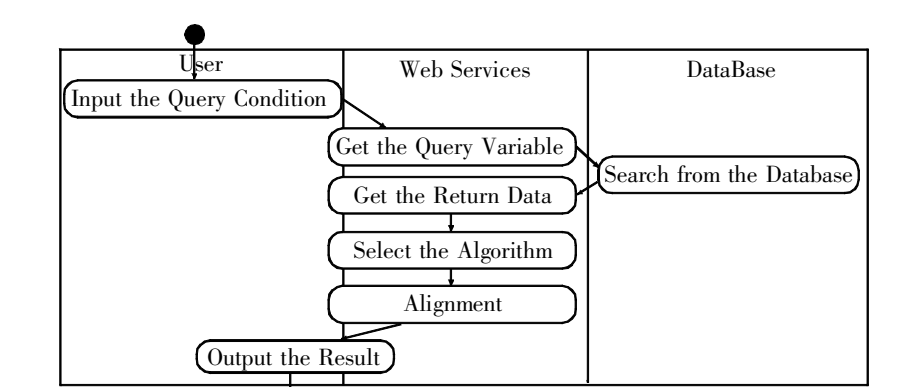


图 5 序列比对流程图

### 3.2 序列比对 Web Service 的设计

由于 .NET 平台的便利性,因此序列比对的例子采用该平台进行程序设计实现。

#### (1) 创建 Web Services 的实例——Alignment 服务

```
Imports System.Web.Services
<WebService(Namespace:= "http://tempuri.org/") >
Public Class Alignment
Inherits System.Web.Services.WebService
End Class
```

#### (2) 构造双序列比对的各种方法

```
Public Property GetSequenceA() As String 获取序列 A
Set( ByVal Value As String)
SequenceA = Value
End Set
End Property

Public Function SmithWaterman( SeqA as String, SeqB as String) As String
使用 SmithWaterman 算法对 SeqA, SeqB 进行序列比对,返回比对结果
End Function
```

### 3.3 双序列比对 Web Service 的调用

对于已建立的 Web Service,其他 Web 应用可以对其引用。比如使用 HTTP-GET,参数同查询字符串一同被传递:

```
http://MyWebSite/Alignment.asmx/SmithWaterman? SeqA = AGCTAGCT&SeqB = AGCT
```

用户还可用 HTTP-POST 方法的参数作为表单值在 POST 体内传递,或者可用 HTTP-SOAP 方法的参数被包裹在一个工业标准 XML 格式中。当然,也可以直接在程序里引用,方法是:首先在 Visual Studio .NET 里面新建一个 ASP.NET Web 的应用程序,然后在引用里,添加“Alignment Web Services”项目。以下是调用 Alignment Web Service 的关键代码:

```
Dim Align As Alignment WebService
Align = New AlignmentWebService()
Align.GetSequenceA = "TAAAAGCTAAAAGCTAAAGCTAGCT"
Align.GetSequenceB = "TAAAAGCTAAAAGCTAAAAAAA"
Print Align.SmithWaterman(SequenceA, SequenceB)
```

由以上方案的实现可知,Web 应用作为服务请求者,看到的只是 Alignment Web Services 所提供的接口函数,对于 Alignment Web Services 内部的规则和数据是完全不知道的。但是用户的 Web Services 应用能很方便地调用此双序列比对 Web Services。以上代码的运行结果略。

## 4 小结

在分析传统的本地和远程生物信息系统模型(下转第 164 页)

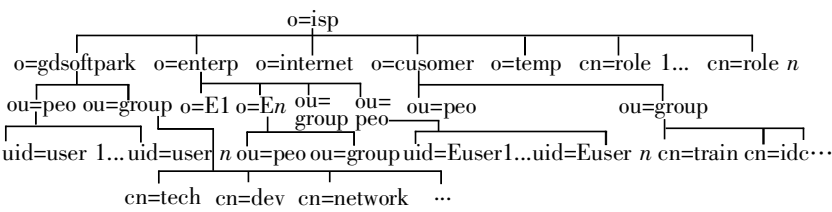


图 3 目录树结构图

3.3 采用 Web Services 实现跨越不同 Web 应用的用户统一身份认证

在对资源进行整合时需要跨越不同的 Web 应用边界,利用不同的服务进行业务间协作。大多数 B/S 应用系统都有自己独立的用户管理和权限管理,即有自己独立的用户认证和授权方式。不同的 Web 应用分别代表分布式环境中不同的独立的安全域,在分布式环境中任意两个安全域间是不存在任何信任关系的,因此,为实现统一的用户身份认证需要在不同安全域间建立信任关系。

整合具有独立的用户认证和授权方式的的应用的过程中需要充分考虑用户关系数据的映射和同步,解决认证和授权结果在安全域间标准化的表示、共享和传递的问题,从而实现安全域间的单点登录机制。具体来说,实现统一身份认证机制需要: 建立用户角色映射机制,即将门户的用户角色和整合的系统的用户类别对应; 在门户和整合的应用系统之间建立数据通信机制,传递用户信息到整合的应用系统进行应用系统自身的授权认证。需要传递的用户信息内容有: 用户注册信息、用户登录确认信息、用户角色确认信息、用户访问资源标志等。

实现跨平台应用的统一身份认证具体实现用户注册、登录、注销等操作时,在门户和集成的应用系统中进行的身份存储和认证过程如下:

- (1) 用户注册。用户注册时将用户信息存储到 LDAP 目录内对应的位置,并由管理员进行身份确认和相应的授权操作。
- (2) 用户登录。用户登录时由门户进行单点登录认证,形成 AuthenticationToken; 其他应用系统可获得用户登录信息及标志,依照用户登录信息缓存策略,检查用户是否已经登录。如果没有,则重定向到门户登录页面进行登录,处理完用户登录后,将重定向到资源的 ReturnURL,并将 AuthenticationToken 传回。
- (3) 用户注销。用户注销时需从其访问过的所有资源中一一注销,用户在任何整合的系统中进行注销操作都会重定向到门户注销页面,由门户逐一遍历各应用服务并使其失效、用户退出。
- (4) 用户权限确认。检查用户是否已经登录并具有某权限。
- (5) 用户数据更新。门户系统 LDAP 用户信息修改时需要与整合的各系统的用户信息保持同步修改。

具体的身份认证服务以 Web Services 的方式提供,并采用 XML 数据格式,利用 HTTP 协议及 SOAP 协议进行通信。身份认证的实现机制如图 4 所示。

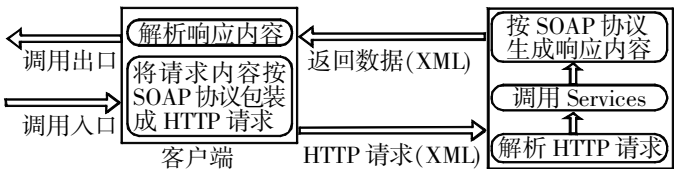


图 4 身份认证的 Web Services 通信机制

3.4 资源配置管理和动态扩展

在公共技术服务平台的建设过程中,资源会不断地引进和更新,因此资源整合系统需实现资源的配置、管理和动态扩充。资源配置管理包括资源的增加、删减,资源的授权分配、访问控制等,需定义资源描述模型和资源访问公共接口。

- (1) 资源描述。主要为资源信息和权限信息,如资源名称、资源类型、访问类型、访问地址、访问限制等。
- (2) 资源访问公共接口。定义了一个统一的公共访问接口,用于公共技术资源平台的各子系统访问资源配置信息。

4 结束语

采用门户、Portlet、Web Services 等技术开发的资源整合系统为用户访问公共技术服务平台的资源提供了统一的入口,用户可通过该系统进行在线申请、注册、使用平台的资源,查询、监控所使用资源的运行状况,同时通过该系统的资源管理模块可进行资源的动态管理和配置,较好地解决了资源平台建设中的动态扩展问题。下一步除了继续完善资源动态配置和访问接口外,还需通过对集成方式和集成技术的进一步研究增加新集成的资源的集成度,从而使用户可更方便、更有效地使用公共技术服务平台的各类资源。

参考文献:

[ 1 ] 美国太阳微系统公司. Sun ONE Portal Server 开发指南[ M ]. 王永茂. 北京: 机械工业出版社, 2003. 1-23.

[ 2 ] 美国太阳微系统公司. Sun ONE Application Server 开发指南[ M ]. 师炜. 北京: 机械工业出版社, 2003. 2-25.

[ 3 ] 美国太阳微系统公司. Sun ONE Directory Server 5.2 管理指南[ EB/OL ]. <http://docs.sun.com/app/docs/doc/816-6852-10?l=zh>, 2003.

[ 4 ] 美国太阳微系统公司. Sun ONE Directory Server 5.2 Reference Manual[ EB/OL ]. <http://docs.sun.com/app/docs/doc/816-6699-10?l=zh>, 2003.

作者简介:

王萍(1967-),女,湖北荆州人,工程师,硕士,主要研究方向为计算机应用、软件工程;李其均(1964-),男,湖北仙桃人,研究员,学士,主要研究方向为计算机系统软件、系统集成技术。

(上接第 158 页)后,根据 Web Services 体系结构思想提出了一个生物信息系统功能服务的模型,此模型可以解决传统生物信息系统的缺点。利用该模型举例说明了一个生物信息系统的双序列比对功能的 Web Services 的实现。我们模型的优点体现在:它使得计算机学家专注于算法设计,生物学家专注于生物数据的分析。毫无疑问,这就是 Web Services 的优越性。

参考文献:

[ 1 ] 柴晓路,梁宇奇. Web Services 技术、架构和应用[ M ]. 北京: 电子工业出版社, 2003. 1-9.

[ 2 ] 北京大学生物信息中心[ EB/OL ]. <http://www.cbi.pku.edu.cn/chinese/>.

[ 3 ] W3C Working Draft. Web Services Architecture[ EB/OL ]. <http://www.w3.org/TR/2004/NOTE-ws-arch-20040211/>, 2004-03-30.

[ 4 ] Richard Anderson. Professional ASP. NET 1.0[ M ]. 北京: 清华大学出版社, 2002. 883-926.

[ 5 ] Microsoft Corporation. Microsoft MSDN Library [ Z ]. 2003.

作者简介:

林毅申(1980-),男,广东广州人,硕士研究生,主要研究方向为生物信息学、信息安全;林丕源(1963-),男,四川蓬安人,教授,硕士生导师,主要研究方向为图形图像处理、信息安全。