

基于网络服务的数字图书馆集成技术¹

王昉(四川大学信息管理系, 成都 610064)

张晓林(中国科学院文献情报中心, 北京 100080)

文摘: 本文提出了基于集成网络服务的数字图书馆的概念模型, 并分析了实现该模型的技术机制—网络服务概念框架和现有标准与技术, 包括 SOAP、WSDL、UDDI、WSFL。

关键词: 网络服务, 数字图书馆, 网络服务概念框架,

分类号: G25, 35

Integration Technology of Web Services-based Digital Library

Wang Fang, Department of Information Management, Sichuan University, Chengdu, 610064

Zhang Xiaolin, Library, CAS, Beijing, 100080

Abstract: This paper presents a conceptual model for Web-Services-based integration of digital libraries, and analyses the Web Services framework as the implementation base and the underlying techniques, SOAP, WSDL, UDDI, WSFL, etc.

Keywords: Web Services, digital libraries, Web Services Conceptual Architecture

1. 前言

随着网络信息资源形态的丰富和用户需求的多样化, 我们需要一种新的机制能将现有的各种网络化分布式信息资源加以组织、集成并以个性化的网络服务的形式提供给用户; 同时, 数字图书馆从过去以资源为中心逐渐演变为以用户信息活动为中心, 将信息资源和信息服务组织、集成、嵌入到用户信息活动和用户信息系统中, 建立个性化的数字图书馆^[1]。本文将就这种基于集成网络服务的数字图书馆的机制及其现有的标准、技术进行初步的探讨。

2. 基于集成网络服务的数字图书馆机制

根据数字图书馆集成要求, 我们分析基于网络服务的需求、模式及其关键任务:

2.1 需求分析

在网络化、信息化的过程中, 全世界的各个数字图书馆、内容提供商、企业及各种机构在不同的操作平台、硬件环境下, 用不同的标准、对象模型和编程语言建立起了各自的信息资源及应用程序。为了实现资源共享, 需要一种简单、易行、低成本的方式, 将现有的信息资源和应用程序封装起来, 以网络服务(Web Service)的形式提供给用户, 并以一种标准的方法描述、发现和集成, 以实现异构系统之间的互操作和动态集成, 从而提供新的应用和服务。这种网络服务应该具备以下特征:

- (1) 独立于系统平台、语言、协议和对象模型, 支持网络服务之间的直接交流;
- (2) 使用标准的方法描述和发现网络服务, 支持用户、内容提供者和合作伙伴所拥有的网络服务之间的直接访问和交流;
- (3) 使用标准的方法实现对远程网络服务的访问和动态绑定;
- (4) 数字图书馆的网络服务器能根据用户信息提供个性化定制服务, 支持对网络服务的自动发现和选择, 很少或不需要人工干预;
- (5) 支持各类 Internet 设备用户(如 PC 接入、无线接入等)使用各种协议(如 HTTP 协议)或代理或网关(如 WAP 网关或语音网关)访问网络服务的服务器; 支持各种设备使用的不同的标记语言(如 HTML、WML、cHTML、VoiceXML 等)。

¹ 本文得到“中国科学院百人计划基金”资助

2.2 基于集成网络服务的数字图书馆模式

根据对数字图书馆的开放性、动态集成、互操作的要求, 集成网络服务体系可包括网络服务、网络服务登记机构和数字图书馆网络服务服务器三个主要组成部分和网络服务的发布、发现、和动态访问规则, 通过一系列转换、整合、服务调用和管理机制^[1]形成为用户提供集成和定制集成网络服务体系。图 1 给出了一个集成网络服务的数字图书馆概念结构。

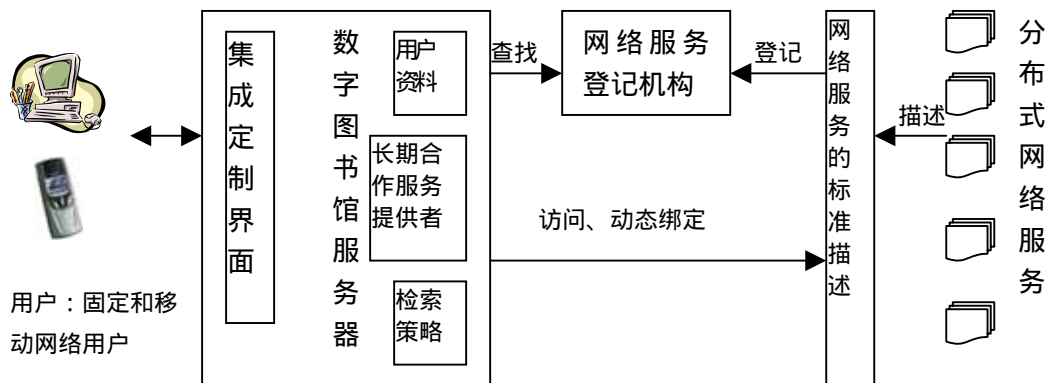


图 1 集成网络的数字图书馆概念模型

3 网络服务集成框架

3.1 网络服务的概念

网络服务 (Web Services) 提供一个标准接口, 用于描述一系列可通过使用标准 XML 传送方式实现网络访问的操作的集合。接口隐藏了服务执行的细节, 允许服务的使用独立于执行它的硬件和软件平台和编程语言。这个特性允许和鼓励基于网络服务的应用之间松散耦合, 面向组件、跨技术应用。网络服务可以完成一项或一系列任务, 可以单独与其他网络服务一起完成复杂的集聚和应用。

3.2 网络服务模型^[2]

网络服务模型 (Web Services Model) 包括三个角色: 服务提供者、服务登记机构和 服务请求者; 三种操作: 发布、查找和绑定; 以及操作的基础: 网络服务软件和服务描述。图 2 给出了网络服务模型的结构:

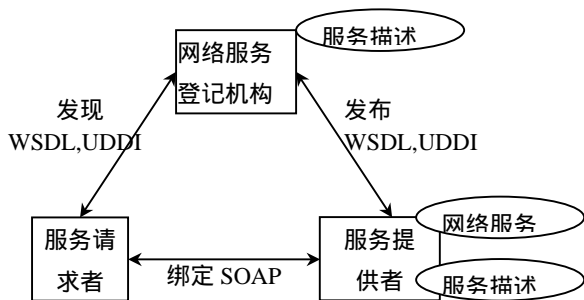


图 2 网络服务的角色、操作

实际运行中, 服务提供者采用一种网络可访问的软件模型作为网络服务的使用方式, 并提供网络服务的描述发布给网络

服务登记机构或服务请求者; 服务请求者通过查找方法检索本地的网络服务描述或查找网络服务登记机构拥有的服务描述, 从而利用服务描述实现与网络服务的绑定和激活/交互。其中服务提供者和服务请求者是可彼此交换角色的逻辑概念。

3.3 网络服务协议栈

为了实现网络服务的发布、查找和绑定的操作, 我们需要一个协议栈在每一层上执行完成不同的任务并纳入相应的标准。图 3 给出了一个网络服务协议栈的逻辑结构图^[3], 其中每一层的功能都建立在它的下一层提供的服务之上; 各层为其上一层提供一个通用的接口, 其内部实现相对其他各层透明。图中还给出了每一层应用的标准技术和需要的服务。

3.3.1 网络层

网络层是网络服务协议栈的基础。它的功能是确保网络服务的网络可达性。网络服务可位于公用 Internet(外联网)上和专用 Intranet(内联网)上,支持它们的网络协议包括 HTTP、FTP、SMTP, Remote Method Invocation(RIM, 远程方法调用)等,使用最普遍的是 HTTP。网络协议的选择除了通用外,还可根据系统的安全性、可获得性、性能和可靠性的需要考虑。大型机构由于复杂的网络结构采用了多种网络协议的情况下,可采用 HTTP 实现各种协议的桥接。

3.3.2 基于 XML 的传递层

这一层的主要功能是应用之间的消息传递,以实现网络服务之间的交流。“消息传递”中“消息”的语义着重于服务而非网络技术,偏重于用文档描述信息。它使用了以 XML 为基础的消息传递技术,并选用 SOAP (Simple Object Access Protocol, 简单对象访问协议)^[4]作为具体的 XML 消息传递协议。SOAP 提供了一个简单、短小的以 XML 为基础的交换网络应用间的结构化数据的机制,自身不定义任何应用的语义例如编程模型或使用特定的语义,而通过提供一个组件封装模型和模型内部数据的编码机制来表达应用的语义。通过这种机制,它可与其它网络协议如 HTTP、SMTP、FTP 等联合使用或被其他协议重新封装使用。大多数网络服务的开发者并不需要直接和 SOAP 接触,网络服务使用的特定编程语言绑定可以由 WSDL 生成。

3.3.3 服务描述层

网络服务描述是网络服务提供者了解激活网络服务所有规定以提供给服务请求者的方法。它是网络服务框架能够独立于具体网络服务的软硬件环境、编程环境等松散的耦合的关键。根据对网络服务描述的深度我们可分为两部分:一是网络服务的基本描述即网络服务的接口和网络服务的执行方法的描述,以支持网络服务之间在技术意义上的互操作;二是网络服务的完整描述,包括各种应用的使用条件、服务质量、服务与服务之间的相互关系等,以支持网络服务之间、服务提供者与服务请求者之间和他们三者之间的在服务意义上的互操作。

基本服务描述包括对网络服务接口的描述和服务执行方法的描述,它们都基于 XML 消息传递。WSDL (Web Services Description Language, 网络服务描述语言)^[5-6]是基于 XML 服务描述的事实标准,它描述了网络服务的接口和交互机制。

完整服务描述建立在 WSDL 描述的基本服务描述之上,包括服务使用的具体环境,如服务的拥有者、类型、产品、相关类目、质量、收费情况等等。UDDI (Universal Description, Discovery and Integration specification, 统一描述、发现和集成规范)^[7]提供了完成这些服务描述的机制。另外,完整的描述还应该包括对多个服务合作者之间关系和流程的描述。现有的网络标准如 ebXML 合作协议纲要 (ebXML Collaboration-Protocol Profile and Agreement Specification)^[8]描述了各个应用合作伙伴之间角色、任务等概念。其它基于网络服务的标准如 WSFL (Web Services Flow Language, 网络服务流程语言)^[9]正在开发之中。

3.3.4 服务发布层

网络服务发布是将网络服务描述的结果提交给网络服务登记机构或服务请求者。根据网络服务被使用的动态程度,可将网络服务的发布方式分为三种:

- 直接发布,服务提供者直接将网络服务描述发送给服务请求者,可以通过电子邮件附件、FTP 甚至光盘。这种情况通常发生在双方已经议定好服务条款、费用已付等情况下。
- 使用 DISCO (Discovery of Web Services, 网络服务发现,微软曾提出的方案)^[10]和 ADS (Advertisement and Discovery of Services, 服务广告和发现, IBM 标准)^[11]。DISCO

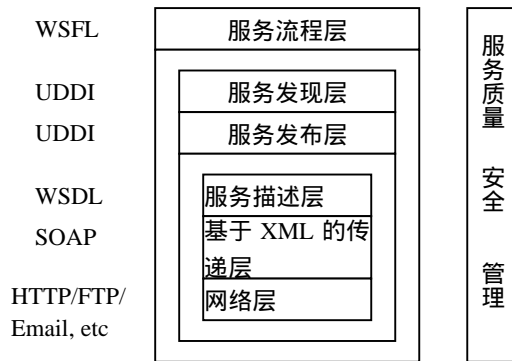


图3 网络服务协议栈

和 ADS 都定义了一个简单的 HTTP GET 机制从给定的 URL 检索网络服务描述, 通常一个增强的网络服务描述存储机构都会有一个附加的检索网络服务的功能;

c. 将网络服务描述发布到一个 UDDI^[7]服务登记机构。可能要求网络服务支持特定标准、检索型元数据、接口或数据类型, 过滤非法条目和要求服务质量的认证。

3.3.5 服务发现层

网络服务发现包括获取服务描述和消费服务。

像发布服务一样, 获取服务描述的方法随服务描述发布的机制和动态程度不同而不同。对于直接发布服务描述方式, 服务描述位于请求者本地的一个文件或一个简单的描述存储器里, 可直接检索; 对于另外两种发布方式, 需要一种查询机制提供通过接口类型(基于 WSDL 模板) 所绑定协议的信息、查询参数、所需中介类型、服务分类、服务提供者信息等的发现。获取服务描述之后, 请求者需要用它激活服务。请求者根据 WSDL 描述的服务接口通过 SOAP 代理(或其他与 WSDL 文档绑定的编程语言代理) 产生一个 SOAP 请求(或特定编程语言的请求) 给所请求的网络服务, 网络服务响应, 完成对服务的绑定。

3.3.6 服务流程层

网络服务的重要特性是他们可像组件一样集成。一项特定事务通常由许多应用组成, 工作流程作为组合网络服务的基本机制, 可以快速创造新的网络服务。

WSFL^[9]是用于描述网络服务组合的 XML 语言, 解决两类问题: 一是规定了针对特定目标、按照特定过程的特定网络服务集合的描述模式, 典型例子是一个事务处理过程的描述; 一是规定了这个集合的各个网络服务之间的交互模式。同时, WSFL 集合支持事务处理中合作伙伴的广泛交互模式, 特别是层级式交互和点到点的交互。层级式常见于长期、稳定的合作伙伴之间; 而点到点的交互常用于临时、动态合作。

实际应用中, 网络服务还应考虑的问题还包括安全机制、服务质量和可信赖传递、系统和应用的管理等, 限于篇幅, 本文不再赘述。网络服务的实例有 IBM 基于 WebSphere Portal Server^[12]的网络服务, 微软基于 .net 技术的网络服务^[13]等, 主要解决网络的商务应用, 基于数字图书馆的网络服务还有待开发。

参考文献:

- [1] 张晓林. 数字图书馆机制的范式演变及其挑战. 中国图书馆学报, 2001.6
- [2] Web Services Activity. <http://www.w3.org/2002/ws/>
- [3] Heather Kreger, IBM Web Services Conceptual Architecture 1.0
<http://www-4.ibm.com/software/solutions/webservices/pdf/WSCA.pdf>
- [4] Simple Object Access Protocol (SOAP) 1.1, W3C Note 08 May 2000. <http://www.w3.org/TR/SOAP/>
- [5] Web Service Description Requirements, W3C Working Draft 29 April 2002.
<http://www.w3.org/TR/2002/WD-ws-desc-reqs-20020429/>
- [6] Web Services Description Language (WSDL) Version 1.2, W3C Working Draft 9 July 2002
<http://www.w3.org/TR/2002/WD-wsdl12-20020709/>
- [7] Universal Description, Discovery and Identification. <http://www.uddi.org/>
- [8] ebXML Collaboration-Protocol Profile and Agreement Specification. <http://www.ebxml.org/specs/ebCCP.pdf>
- [9] Web Service Flow Language. www-4.ibm.com/software/solutions/webservices/pdf/WSFL.pdf
- [10] Discovery of Web Services. <http://lists.ebxml.org/archives/ebxml-transport/200007/msg00007.html>
- [11] The Advertisement and Discovery of Services (ADS) protocol for Web services
http://www-900.ibm.com/developerWorks/cn/web/ws-ads/index_eng.shtml
- [12] WebSphere Portal Server. <http://www-3.ibm.com/software/webservers/portal/>
- [13] Microsoft .NET. <http://www.microsoft.com/net/>