

基于 Web Services 异构资源汇集与共享系统

李国栋, 李 桦, 柳长安

(华北电力大学 计算机科学与技术学院, 北京 102206)

摘 要:以网络科技资源应用集成环境平台的研究和开发为背景,针对网络科技资源分布广、共享度低、资源信息形式多样、没有统一的标准等现状,对 B/S 模式下异构资源汇集与共享的需求进行了分析,提出了基于 Web Services 的异构资源汇集与共享系统的设计方案并实现。通过该系统可有效的实现国家科技基础条件平台的科技资源汇集,在统一的元数据规范下,对所汇集的科技资源形成统一的存储形式并进行目录管理;使国家各领域平台的科技资源能够快速、准确、高效地被整合利用,以适应国家科技基础条件平台资源应用服务集成的需求,提升网络科技服务系统的技术水平,扩大应用范围,增强服务效能。

关键词:异构资源; Web Services; 浏览器/服务器模式; 资源汇集; 资源共享

中图分类号: TP311

文献标识码: A

文章编号: 1673-629X(2008)12-0137-03

Web Services - Based Solution for Heterogeneous Resources Integrated and Shared System

LI Guo-dong, LI Xi, LIU Chang-an

(School of Computer Science and Technology, North China Electric Power University, Beijing 102206, China)

Abstract: In the background of research and development to the network scientific and technological resources application integrated environment platform, in order to solve the problems, such as wide distribution of network scientific and technological resources, low sharing, various forms of resource information and no uniform standard, a demand analysis for the B/S mode heterogeneous resources integrated and shared has been carried out. The design of heterogeneous resources integrated and shared system which based on the Web services is proposed and realized. Through this system, the collection of the scientific and technological resources of the national basic condition platform can be effectively completed, and could uniform the storage style of the collected scientific resources and implement the inventory management; national platform in the areas of science and technology resources can be integrated and utilized rapidly, accurately and efficiently so as to meet the needs of the basic conditions platform of national scientific and technological resources application service, and to upgrade the technical level of network technology services, to expand the scope of application and enhance efficiency.

Key words: heterogeneous resources; web services; B/S mode; resources integrated; resources shared

0 前 言

随着网络技术在社会各个领域的迅猛发展和互联网上资源的迅速积累,海量数据的共享和异构数据源的统一管理已成为当务之急。“信息孤岛”^[1]现状一直困扰着整个信息现代化建设。

网络科技资源分布广,共享度低;资源信息形式多样,没有统一的标准;科技资源种类繁多,很多科技网站及相关信息资源的分类分级体系存在着不科学、不规范的情况,难以进行整合集成;资源数量极其庞大,

增长迅速。据不完全统计,全互联网提供的科技信息总量超过 20TB^[2],而且正在以每年高于 25% 的速度激增^[3],但利用率很低;资源需求者无法及时、准确地找到所需资源,导致科技资源建设工作进展缓慢;资源需求多样化,传统的门户资源展示已无法满足社会需求。基于 Web Services 的异构资源汇集于共享系统采用 B/S 模式,通过资源汇集机制将资源汇集于“一点”,为国家科技基础条件平台提供统一、高效、安全、可靠的资源访问服务。

1 Web Services 技术

Web Service 可以被认为是一种部署在 Web 上的对象(Web Object),因此,具有对象技术所承诺的所有优点。同时,Web Services 的基石是以 XML 为主的、

收稿日期:2008-04-21

基金项目:国家科技基础条件平台建设资助项目(2005DKA63904)

作者简介:李国栋(1974-),男,硕士,副教授,研究方向为机器人技术、信息处理。

开放的 Web 规范技术,因此,具有比任何现有对象技术更好的开放性。

1.1 Web Services 体系架构

Web Services 体系结构是基于 3 种角色(服务提供者、服务注册中心和服务请求者)之间的交互。图 1^[8]显示了这些操作及提供这些操作的组件及其之间的交互^[4]。

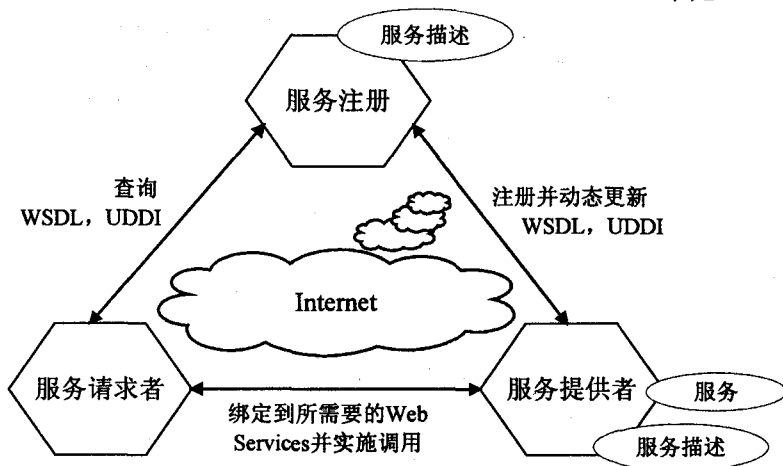


图 1 Web Services 角色、操作和构件

(1)服务提供者。从企业的角度看,这是服务的所有者。从体系结构的角度看,这是托管访问服务的平台。

(2)服务请求者。从体系结构的角度看,这是寻找并调用服务,或启动与服务交互的应用程序。服务请求者角色可以由浏览器来担当,由人或无用户界面的程序(例如另外一个 Web 服务)来控制。

(3)服务注册中心。这是可搜索的服务描述注册中心,服务提供者在此发布他们的服务描述。在静态绑定开发或动态绑定执行期间,服务请求者查找服务并获得服务的绑定信息(在服务描述中)。对于静态绑定的服务请求者,服务注册中心是体系结构中的可选角色,因为服务提供者可以把描述直接发送给服务请求者。

1.2 Web Services 协议

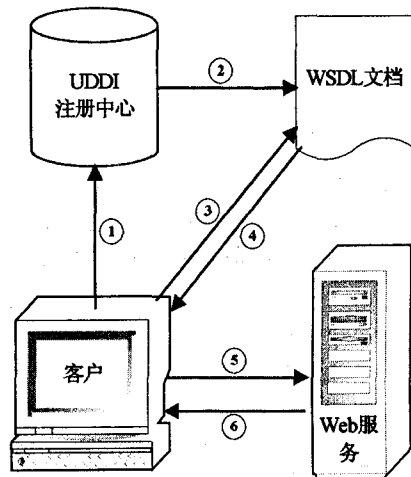
Web Services 区别于其他类似计算模型的部分原因在于 Web Services 使用了 XML 和基于 XML 的标准。最常见的是 SOAP、WSDL、UDDI。这些技术使应用程序之间的通信能够以一种独立于特定编程语言、操作系统和硬件平台的方式进行^[5](见图 2^[6])。

(1)XML 规范。XML(可扩展标记语言)由标准通用标记语言(SGML)发展而来,现已成为一种人们广泛接受的用于描述数据和创建标记语言的标准。同许多由专用走向标准的技术不同,XML 是由 W3C 所定义的一种开放的、标准的技术。

(2)SOAP 协议。SOAP(简单对象访问协议)是一种基于 XML 的、不依赖传输协议的、用于在应用程序之间以对象的形式交换数据的表示层通信协议,是 Web Services 的核心,可以看成是用户端与服务器端之间进行沟通的特殊语言。

(3)WSDL 协议。Web Services 的另外一个重要标准是 WSDL(Web 服务描述语言)。其被用来描述 Web Services 的相关信息,它把网络服务定义成一个能交换消息的通信端点集。

(4)UDDI 协议。UDDI(统一描述、发现和集成协议)是一套基于 Web 的、分布式的、为 Web Services 提供的信息注册中心的实现标准规范,是一个跨产业、跨平台的开放性架构,可以帮助 Web Services 开发商在 Internet 上公布自己推出的 Web Services,同时也可以被企业发现。



注:①客户查询注册中心以找到服务的位置;
②注册中心引导客户找到 WSDL 文档;
③客户访问 WSDL 文档;
④WSDL 提供与 Web 服务进行交互的数据;
⑤客户发送 SOAP 消息请求;
⑥Web 服务返回 SOAP 消息应答。

图 2 Web Services 交互中的 SOAP、UDDI 和 WSDL

2 系统设计

基于 Web Services 异构资源汇集与共享系统的研究目标及意义在于:充分利用各个节点资源优势,利用先进的网络信息技术,汇集国家科技基础条件平台的科技资源,迅速提升资源利用率;解决“信息孤岛”^[1]现状给信息现代化建设带来的一系列严重问题。

2.1 设计方案

1)建立一个中心交互界面,如中心数据库,为异构

资源的存储提供支持;另外在其服务站点上提供 Web Services 接口,用来实现中心数据库与异构网络资源服务器的数据库之间的通信功能,这主要为了实现异构资源远程检索与共享。

2)设置一个基于 Web Services 实例实现的服务,通过适配器在各异构数据库所在局域网的对外服务器上实现代理数据接入网关对异构应用系统的数据库进行访问,以解决无法访问数据库的问题。

3)充分利用 XML 语言的描述能力描述各种配置信息及数据集^[7]。配置信息主要包括元数据、数据结构、中英文描述等信息。

2.2 系统流程图

1)异构资源汇集流程(见图 3);

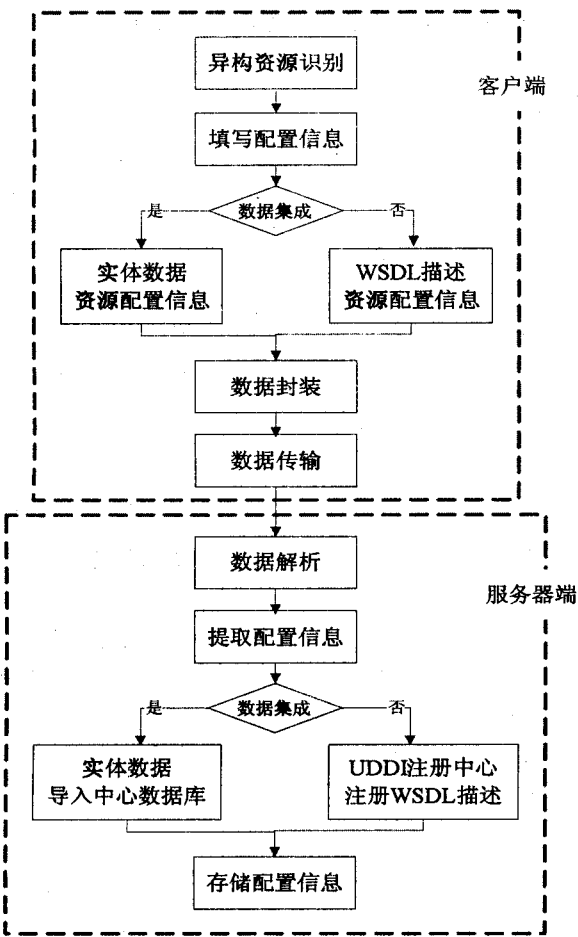


图 3 异构资源汇集

2)异构资源共享流程图(见图 4)。

2.3 关键技术

1)异构资源识别:主要针对面向科技领域的资源信息,是将用户存储在不同类型数据库服务器中的结构相异数据文件,以都柏林核心元数据^[8]为标准,提交到服务器端,目的在于屏蔽各个资源提交节点异构的数据格式,以统一的数据文件格式存储在服务器端,便于数据的整合和汇集。

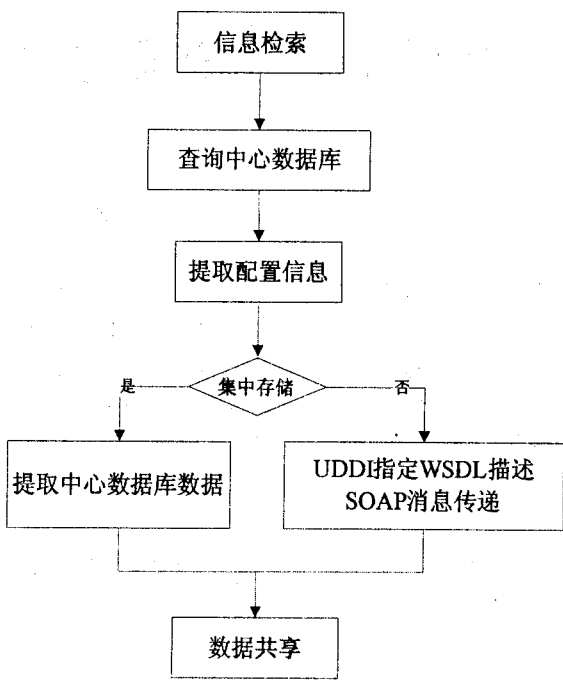


图 4 异构资源共享

2)资源集中存储:资源的异构性导致对资源集中存储存在很大的困难,本系统通过记录异构数据文件基本信息,在服务器端对这些信息进行识别并通过存储过程创建同资源提交端相同的数据表,而中心数据表只记录与其对应的映射信息。

3) Web Services 技术: Web Service 技术克服了 DCOM、CORBA 和 Java RMI 等传统的集成技术所受到的网络环境很大程度上的限制。Web Services 基于 XML 文档进行服务描述、服务请求和反馈结果,可以在 Internet 上通过 HTTP 协议进行传递,很容易地被访问和返回结果,并能突破防火墙的限制,同时由于相关标准都是 W3C 的开放协议,可以使异构平台上的应用集成做到与平台和操作系统无关;此外, Web Services 是一种动态的集成方案,所有的服务都可以通过 UDDI 标准动态地被发现、绑定和使用,容易适应系统的变动,提高系统的灵活性和伸缩性。

3 案例应用

上述设计方案在网络科技资源应用集成环境平台建设中得到应用。该项目旨在通过屏蔽科技资源各领域平台的差异,实现统一的资源传输规范,依据标准服务接口,采用资源聚类、网络资源自动采集等多种资源汇集模式,对国家科技资源进行汇集,使国家各领域平台的科技资源能够快速、准确、有效地被整合利用,以适应国家科技基础条件平台资源应用服务集成的需求,提升网络科技服务系统的技术水平,扩大应用范

(下转第 143 页)

中权值较大的关键词即可作为本体的特征指数。最终得到的本体特征指数形式如下所示:

$$\{Item1/W_{ime1}, Item2/W_{ime2}, \dots, Itemn/W_{imen}\}$$

得到了这组特征指数后,可用来对本体进行标注。在本体搜索时就可根据输入的查询条件与特征指数进行匹配,并按照权值的大小对结果进行排序。

3 结束语

本体是语义 Web 中各种智能应用的基本工具。在当前本体数量越来越多且又没有统一的本体建模标准的情况下,高效的检索本体变的越来越重要。介绍了一种基于描述逻辑产生式算法的通过语义关键词来标记本体的方法。文中的主要贡献在于通过对本体所包含的概念进行拆分和标记,使得本体与一组能够高度概括其语义信息的关键词联系起来,通过计算各关键词的权值,选出本体中最能体现本体特点的一组关键词,这组关键词就成为该本体的特征指数。

利用描述逻辑中 Tableaux 算法进行概念拆分的过程是比较耗时的,它的时间复杂度最坏的情况下是多项式级别的。但这并不影响该方法的可用性。

该方法还有待解决的几个问题:

1)对甄选结果的评估。当前的评估只能借助于领域专家来评判。自动评估机制有待进一步的研究;

2)如何通过拆分出来的关键词反向生成本体。当

前的本体构建主要是手工或半自动构建。文中提出的方法如何用来指导本体构建和设计也是下一步研究工作的重点。

参考文献:

- [1] Berners-Lee T, Hendler J, Lassila O. The Semantic Web [J]. Scientific American, 2001, 284(5): 28-37.
- [2] Dean M, Schreiber G, van Harmelen F, et al. OWL Web Ontology Language Reference [EB/OL]. 2003-08-18. <http://www.w3.org/TR/owl-ref/>.
- [3] Hu B, Dasmahapatra S, Dupplaw D, et al. Reflections on a medical ontology [J]. International Journal of Human-Computer Studies, 2007, 65(7): 23-25.
- [4] Ding L, Pan R, Finin T, et al. Search on the semantic Web [J]. IEEE Computer, 2005, 10(38): 62-69.
- [5] RDF Core Working Group, RDF Reference [EB/OL]. 2004. <http://www.w3.org/RDF>.
- [6] 史忠植,董明楷,蒋运承,等.语义 Web 的逻辑基础 [J]. 中国科学 E 辑. 信息科学, 2004, 34(10): 1123-1138.
- [7] Baader F, Calvanese D, McGuinness D, et al. The Description Logic Handbook: Theory, Implementation and Applications [M]. Cambridge: Cambridge University Press, 2003.
- [8] Horrocks I, Sattler U. A tableaux decision procedure for SHOIQ [C] // In Proc. of the 19th Int. Joint Conf. on Artificial Intelligence (IJCAI 2005). [s. l.]: Morgan Kaufman, 2005.

(上接第 139 页)

围,增强服务效能。

基于 Web Services 异构资源汇集与共享系统很好地实现了该项目的宗旨,通过客户端软件将异构资源进行标准化处理,在服务器端获取数据资源和配置信息进而存储,屏蔽了资源的异构性;用户通过网络科技资源应用集成环境平台门户网站检索需要的资源,使用户对资源的访问透明化。

4 结束语

文中是在参与国家科技基础条件平台应用服务支撑系统的子项目“网络科技资源应用集成环境平台建设”的过程中,针对网络科技资源分布广泛、资源共享程度不高、资源信息形式多样、种类繁多等特点,利用 Web Services 技术,按照面向对象的设计方法,设计并实现了基于 Web Services 异构资源汇集与共享系统。该系统能有效地屏蔽科技资源各领域平台的差异,实现统一的资源传输规范,使国家各领域平台的科技资源能够快速、准确、有效地被整合利用,目前该系统已上线试运行。

参考文献:

- [1] 马大川,杨红平.信息资源的集成整合研究 [J]. 中国图书馆学报, 2004, 30(3): 36-40.
- [2] Li Jun, Furuse K, Yamaguchi K. Focused Crawling by Exploiting Anchor Text Using Decision Tree [C] // Proceedings of the 14th International World Wide Web Conference. Chiba, Japan: [s. n.], 2005: 1190-1191.
- [3] Cheng Jing, Li Qing, Wang Li ping, et al. Automatically generating an e-textbook on the Web [C] // In: Lecture Notes in Computer Science. Berlin: Springer-Verlag, 2004: 35-42.
- [4] 柴晓璐,梁宇奇. Web Services 技术、架构和应用 [M]. 北京:电子工业出版社, 2003.
- [5] Deitel H M, Deitel P J, Duwaldt B, et al. web 服务实用技术教程 [M]. 北京:机械工业出版社, 2004.
- [6] 屈良. 基于 Web Services 的网络信息资源集成研究 [J]. 中国信息导报, 2006(9): 56-61.
- [7] 林彤,舒真才. 基于 web Services 的异地异构数据库的集成 [J]. 北京工业大学学报, 2005, 31(2): 210-213.
- [8] 吴建中. DC 元数据 [M]. 上海:上海科学技术文献出版社, 2000.