

面向服务构架的海洋数据集成系统的设计与实现^{*}

郭忠文, 尚传进, 管恩花

(中国海洋大学 计算机系, 山东 青岛 266071)

摘 要: 以 SOA 架构和 Web Services 技术为支撑点, 设计实现了一个海洋数据集成系统, 针对海洋学科中数据源的异构性问题, 采用 XML 技术进行集成, 实现了海洋数据的统一表示。与传统架构的数据集成系统相比, 该系统具有易于扩充、服务友好等优点。

关键词: SOA; 海洋数据集成; XML; 参数字典

中图法分类号: TP319 文献标识码: A 文章编号: 1001-3695(2006)02-0151-04

A Marine Data Integration System Based on SOA

GUO Zhong-wen, SHANG Chuan-jin, GUAN En-hua

(Dept. of Computer Science, Ocean University of China, Qingdao Shandong 266071, China)

Abstract: Discuss the systematic structure and implement details of a marine data integration system, which is based on SOA and web services. Considering the heterogeneity of marine data sources, XML technique is applied to integrate them and their uniform representation is achieved. Comparing with systems using traditional structure, this one is much easier to be extended and has more friendly services.

Key words: SOA; Marine Data Integration; XML; Parameter Dictionary

1 引言

(1) 海洋数据的特点与标准化

海洋信息交换领域一直处于这样一种形势:大量的海洋学资料被越来越多的机构和不同研究目的的科学家收集,而且国际上对建立区域和全球数据库,实现数据资源共享的呼声也越来越高。然而,海洋学科内容丰富,涉及到海洋物理、海洋生物、海洋化学、海洋气象、海洋经济、海岸带等许多研究领域,不同的领域数据采集的设备不同,信息处理的平台不同,数据存储的格式也不同,致使数据很难实现交换和共享。联合国教科文组织(UNESCO)的国际海洋资料交换委员会(IODE),自成立四十多年来,一直致力于海洋数据格式标准化研究,旨在简化数据交换,推动全球海洋技术的发展,但由于种种原因未能取得成功。自1998年2月,Web技术标准化集团W3C(World Web Consortium)推出可扩展标记语言XML(eXtensible Markup Language)以来,国际上一些海洋强国及研究机构逐渐意识到XML技术在处理多种格式的海洋数据、简化数据交换方面存在着巨大潜力。

目前,国际上两大海洋XML研究组织SGXML(ICES-IOC Study Group on XML)和EU Marine XML在海洋数据格式标准化制定方面已取得很大进展。海洋XML标准制定主要集中于三个领域:参数字典开发、海洋数据描述、元数据。现已完成多个参数字典之间映射的XML Schema,英国海洋数据中心(British Ocean Data Center, BODC)参数字典也已得到进一步完善^[7];"Keeley Bricks"理论实现了多种仪器测量的海洋数据XML统一;关于元数据,各研究组织一直在对已有标准如ISO-19115, EDMED等进行比较、优化,希望作出最优的选择^[8,9]。

XML通用的一个重要原因是它具有强大的数据描述能力。XML是一种"原语言",也就是定义语言的语言,它采用一种结构化的数据格式,可以存储复杂的数据。XML文档的内容与格式分离,易于被机器读取和处理,通过一个XML文档就可以把数据信息的内容、结构和含义完整地表达出来,可以用文档类型定义(Document Type Definition, DTD)或者XML Schema定义XML文档的结构,而使用可扩展样式语言(eXtensible Stylesheet Language, XSL)来指定XML文档的显示样式。目前,XML可以作为一种通用格式在系统之间进行数据传递,因此XML已成为Internet环境下事实上的数据表示和数据交换的标准。

然而,尽管美国、澳大利亚、俄罗斯以及欧洲的其他海洋强国都已切实有效地开展了海洋XML的研究应用和开发工作^[1],利用传统的软件构架设计的数据交换系统仍然存在着耦合度高、灵活性差以及无法很好地适应用户数据请求的多样性等问题。本文以Web Services和SOA为技术支撑,设计开发了一个海洋数据集成交换系统,并通过试运行初步证明了系统的有效性。

(2) 面向服务构架(SOA)及其特点

Internet的飞速发展使得当今计算机系统的以下几个特点日益突出:大量异构系统并存,计算机硬件工作方式不同,操作系统不同、编程语言也不同;大量、频繁的数据传输仍然速度缓慢并且不稳定;版本升级无法完成,根本无法知道互联网上有哪些机器直接或者间接地使用某个服务。传统的面向

对象和面向组件的构架已经无法很好地满足当前计算机系统的需求,尤其是无法解决在 Internet 环境下的不同商业应用之间的业务集成问题,面向服务的构架(Service-Oriented Architecture, SOA) 便应运而生,并很快表现出强劲的生命力。P. Baglietto 等人以 XML 和 Web Services 为基础,讨论了商业领域 SOA 系统的基本配置,并给出了热那亚进出口公司货运管理系统的实现方案^[4]。在政府办公和电子政务方面,SOA 也显示出了突出的优势,M. Marchese 以 SOA 为基础,设计了电子政务系统,极好地满足了广大市民对于政府电子服务的多样性需求,得到了广泛好评^[5]。

SOA 不是一种语言,也不是一种具体的技术而是一种软件系统架构,其设计目标是以服务为基础,实现软件系统不同部分的松耦合^[3]。与传统的系统架构相比,它具有如下特点:

- 粗粒度、松耦合的服务构架;
- SOA 以服务为基础,能灵活地适应服务的多变性;
- 具有精确定义的标准化接口。

服务是 SOA 的生命线,任意两个软件实体,如果一方可以为另一方完成一项工作并提供结果,此时就称一方为另一方提供了服务,提供服务的一方称为服务提供者,被服务的一方称为服务消费者。服务交互必须是明确定义的,然而这种定义只是描述性的,任何实现的细节都由服务提供者完成。Web Services 是 SOA 实现中最为流行的一种,它使用 Web 服务描述语言(Web Services Description Language, WSDL) 来描述服务,但不包括服务实现的任何技术细节,WSDL 可以由 HTML 或 XML 语言来充当。另外服务应该是独立的、自包含的,在实现上不应该依赖于其他服务的上下文和状态。

2 系统总体结构设计

系统总体结构如图 1 所示,对应系统硬件配置信息如图 2 所示。

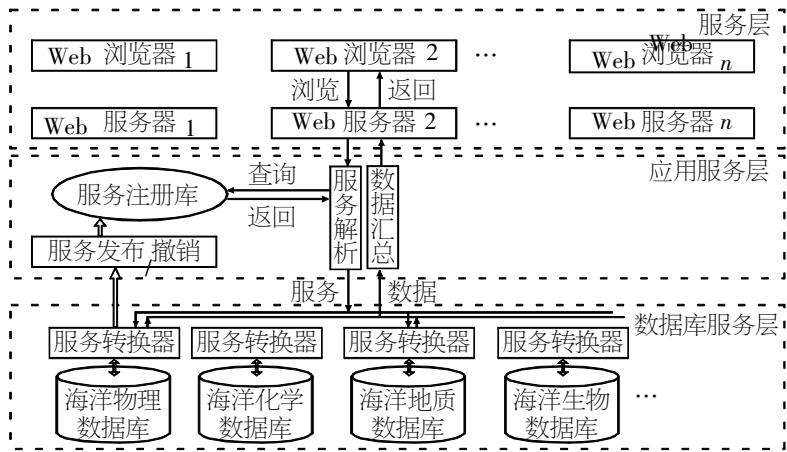


图 1 系统总体结构图

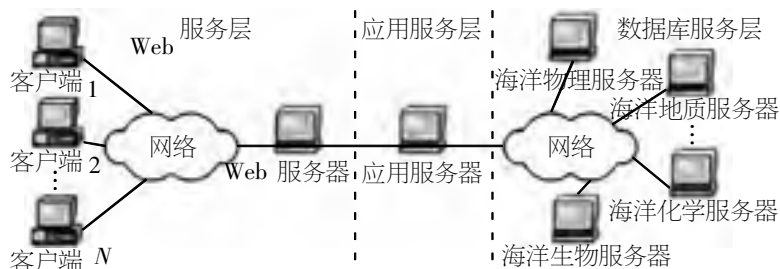


图 2 系统的网络拓扑结构图

2.1 Web 服务层

Web 服务层由分布于各地的 Web 服务器和终端用户的浏览器构成,是终端用户与系统交互的接口。终端用户通过浏览

器以 XML 文档的形式向 Web 服务器发送对海洋信息的 XML 查询请求,并从 Web 服务器接收到 XML 文档格式的查询结果,然后通过 XML Parser,参照定义的 XML DTD 对此 XML 文档进行解析,并以 XML 或 SVG 图像格式在网页上发布。利用 SVG 生成的二维图像可随意缩放、平移,信息实时刷新,不存在页面闪烁问题。

2.2 应用服务层

应用服务层从 Web 服务器接收服务请求,并将服务向能提供服务的数据库服务器转发。本层主要由三部分组成: 服务解析模块。本模块负责从 Web 服务器接收服务请求,并在服务注册库中对该服务进行检索,找到提供该服务的数据库服务器,然后将服务请求向数据库服务器转发。 服务注册库。记录着数据库服务器能提供的服务,当数据库服务层添加新的服务时,需要通过服务发布模块将服务发布到服务注册库中。 数据汇总模块。本模块将数据库服务层返回的数据进行汇总,以 XML 格式返回给 Web 服务器。

2.3 数据库服务层

数据库服务层为系统提供各种海洋信息数据,然而由于学科和采集方法的差异,这些数据无论是在描述还是存储方法上都存在着重大的差异,所以必须为每一个数据库服务器配置一个服务转换器。服务转换器可以将应用服务器提交的服务转换成针对于相应类型数据库服务器的具体操作方法,进而从数据库中获取需要的数据,并将返回的数据转换成统一的 XML 格式数据,提交给应用服务层的数据汇总模块。

3 系统关键技术与实现方案

3.1 数据交换标准制定

数据交换标准在整个系统中起到关键性的作用,是整个系统进行数据交互的基础。本系统遵照以下流程,初步建立起了一个海洋 XML 数据内部交换标准。

(1) 依据 EDMED, CSR, NODC DDF, MEDI/DIF 与 ISO-19115,制定出海洋元数据;参照 BODC 参数字典和海洋各学科 GB 标准,构造参数字典;参照“Keeley Bricks”理论,实现海洋各学科的数据描述。

(2) 在海洋 XML 标准制定过程中充分利用 IODE 的最新海洋 XML 研究成果,总体上符合已有的五个国际标准:

- ISO 19135 - Procedures for Registration of Geographic Items
- ISO 19110 - Feature Type Cataloguing Methodology
- ISO 19126 - UML/XML Implementation of 19110, 19135
- ISO 19136 - GML
- ISO 19139 - GML/19115 Metadata Implementation

(3) 利用所制定标准,整理了中国海洋大学多年的海洋调查数据:全球水深资料(DBDB5) 和西北太平洋水深资料、1990—1999 年全球 SST 产品资料和东海再处理 SST 资料及渤海、黄海、东海、南海的历史海上调查资料等。

3.2 服务的描述

服务注册库以 XML 文件的形式组织(图 3), 每一个子节点对应一个服务登记项,节点各字段含义如下所示:

ServerAddr 为提供服务的服务器地址,以 IP 地址或域名的方式给出。

Port 为提供服务的端口号, 服务转换器采用 C/S 工作模式。
Service 为服务内容, 描述本服务的实际内容。

AvailDate 为服务有效日期, 服务超出有效日期将不再可用, 需要向服务提供者重新确认。

Quality 为服务质量, 是应用服务器根据服务提供者服务速度以及数据质量等参数的综合评价。

Price 为服务价格, 由服务提供者自主定价。

例如节点 Service1 代表的含义为: 服务器 211. 64. 149. 99: 8000 可以 \$5.5 /Kb 的价格提供 “salinity” 数据服务, 服务质量为 “good”, 服务在 01/01/2004 到 12/31/2004 之间有效。

3.3 服务发布与撤销

- SOA 结构中共有三种角色:
- (1) 服务提供者。发布自己的服务, 并且对使用自身服务的请求进行响应。
 - (2) 服务中介。注册已经发布的服务提供者, 对其进行分类, 并提供搜索功能。
 - (3) 服务消费者。利用服务中介查找所需的服务, 然后使用该服务。

本系统中, 数据库服务器充当服务提供者角色, 应用服务器充当服务中介角色, 而终端用户则是服务的消费者。当数据库服务器有新的服务时, 利用 XML 语言对服务进行描述, 并发送到应用服务器, 应用服务器接收到之后将其写入服务注册库中。

3.4 服务转换器设计

表 1 是某数据库服务器的参数字典。
表 1 数据库服务器参数字典实例

参数分类	BODC 中参数名称	数据库服务器中映射名	备注
服务类型	GET	SELECT	查询数据
	MAX / MIN	SELECT-MAX / MIN	统计最大 / 小值
	COUNT	SELECT-COUNT	统计数据个数
	AVERAGE	SELECT-AVERAGE	统计平均值
	...	...	...
时空条件 字段	Time	Measure Time	数据测量时间
	Longitude	Longi	测量点经度值
	Latitude	Lati	测量点纬度值
	Depth	Depth	测量点深度值
	...	...	...
数据信息 字段	Salinity	Salinity	测量点海水盐度值
	Temperature	Temperature	测量点海水温度值
	...	...	...

服务转换器是运行在数据库服务器上的一段程序, 它接收应用服务器发来的服务请求, 将服务解析为针对本服务器数据格式的具体操作并执行(例如当数据存放在标准数据库表中时, 则生成标准的 SQL 语句; 当数据以 XML 文档形式存放时, 则生成 XML 查询语言 Xquery 语句), 将返回的数据按照引言中(1)所述标准格式化为标准的 XML 格式数据, 提交给应用服务器。

为了完成服务的翻译工作, 必须为每一个服务转换器设置参数词典, 需要翻译的参数可以分为两类, 即服务类型和数据库字段。服务类型 = { GET, MAX, MIN, COUNT, AVERAGE, ... }, 数据库字段主要有两类, 即时空条件字段和数据信息存放字段。在服务中使用的这些字段的名称以 BODC 定义的参数字典为准^[7], 数据库服务器参照这一标准, 建立起自己的参数字典, 完成 BODC 标准参数与数据库数据字段的映射。

4 一次服务请求实例

现假定用户 A 需要获取海水盐度数据, 地域范围是: 东经 120 到东经 140 ; 北纬 30 到北纬 45 ; 当前已有海洋物理和海洋地质数据库可以提供海水盐度服务, 服务过程如下所示:

(1) 用户提交服务请求。终端用户将请求以如图 4 所示的 XML 格式描述。

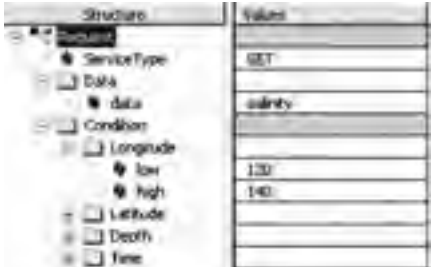
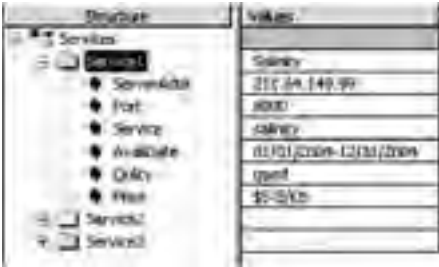


图 3 服务注册库中服务的描述 图 4 用 XML 描述的服务请求

- (2) 应用服务器查找服务提供者, 并向提供者转发服务请求。本例中可以查询到海洋物理和海洋地质数据库两个服务提供者。
- (3) 数据库服务器上运行的服务转换器解析服务并执行, 将执行结果以 XML 格式返回应用服务器。假设数据库服务器的参数字典如表 1 所示, 以下是数据分别存放在标准数据库表和 XML 文件时的服务响应过程:

如果数据存放在标准的数据库表中(假设表名为 SanityData), 服务转换器首先解析服务类型, “GET”应当解析为普通的 “SELECT” 语句, 其次解析欲获取的数据信息字段, 查询参数字典得知 “Salinity” 对应数据库中字段为 “Salinity”, 然后解析时空条件, 例如 “Longitude” 对应 “Longi” 字段, 上下限分别由 “Low” 和 “High” 指定, 最终得到标准 SQL 语句: SELECT Salinity from SanityData where (Longi between 120 and 140) and (Lati between 30 and 45)。执行 SQL 语句, 并将返回结果标准化为标准的 XML 数据, 发送给应用服务器。

如果数据以 XML 文件格式存放(如澳大利亚数据中心就采用图 5 所示结构的 XML 文件对海洋数据进行描述^[6]), 服务转换器将不能使用标准的 SQL 语句获取数据, 必须生成针对 XML 文件的专有操作方式。如图 5 所示, SpatialReference 节点描述了测量点的经纬度, TemporalReference 节点描述了测量值的测量时间, 数据存放于节点 Data 中, 所以根据服务请求可以生成查询条件: //MarineDataRecord/SpatialReference/GeoPoint/Coordinates[@ Latitude > = 30 and @ Latitude < = 45 and @ Longitude > = 120 and @ Longitude < = 140]。利用 ASP 执行当前服务如下所示:

```
<%
    获取东经 120 °到东经 140 °, 北纬 30 °到北纬 45 °的测量数据
    strSourceFile = Server. MapPath( " data. xml" ) 假设文件名为 da-
    ta. xml
    Set objXML = Server. CreateObject( " Microsoft. FreeThreadedXML-
    DOM" )
    objXML. load( strSourceFile)
    Set nodes = objXML. documentElement. selectNodes( " //MarineDa-
    taRecord/SpatialReference/GeoPoint/Coordinates [ @
    Latitude > = 30 and @ Latitude < = 45 and @ Longitude
    > = 120 and @ Longitude < = 140] " ) 指定节点
    循环取得数据
    for i = 0 to nodes. count-1
        nodes( i). childNodes( " Data"). childNodes( " DataObject").
        childNodes( " ValueList"). Text
    next
    set objXML = nothing
%>
```

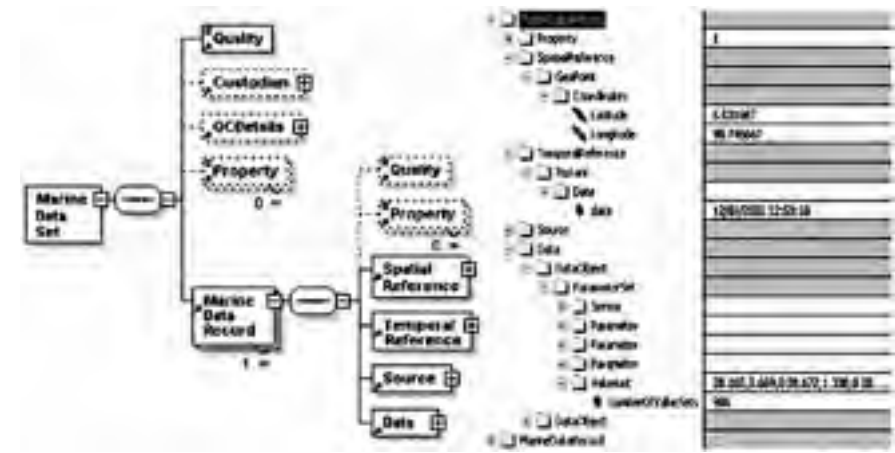


图 5 XML 描述的海洋数据结构示意图

(4) 应用服务器将返回的所有数据汇总, 返回给服务请求者。

可以看出, 与传统构架的数据集成系统相比, 本系统具有两个优点: 数据请求以 XML 描述的服务为基础, 对终端用户具有统一的接口; 系统耦合度低, 数据源的变动只需在应用服务器中更新其服务注册信息, 整个软件系统不需做任何改动, 降低系统升级成本。

5 总结

利用传统的软件构架实现的共享平台耦合性大, 添加新的数据源将带来软件系统的重大调整, 然而在海洋信息交换领域, 增加新数据源是不可避免的, 传统的软件构架很难适用。本文利用 XML 强大的数据描述能力, 制定了兼容国际标准的海洋 XML 数据内部交换标准, 将海量的异构海洋数据进行了统一描述, 成功设计了一套 SOA 构架的海洋数据集成系统, 该系统耦合度低, 能很好地适应海洋数据源的变动。相信随着 XML 技术和 SOA 架构技术研究的进一步深入, 本系统将得到

进一步的完善, 进而大大推动海洋数据服务领域的发展。

参考文献:

[1] 陈继香. 在海洋数据服务领域的应用研究 XML[J] . 海洋通报, 2004, 23(4) : 46-50.

[2] Tonill Hamre, Keiran Millard, Brian Matthews. Analysis of Marine and XML Data Standards[EB/OL] . <http://www.marinexml.net/>.

[3] What Is Service-Oriented Architecture[EB/OL] . <http://webservices.xml.com/pub/a/ws/2003/09/30/soa.html>.

[4] P Baglietto, M Maresca, A Parodi, *et al.* Deployment of Service Oriented Architecture for a Business Community[C] . The 6th International Enterprise Distributed Object Computing Conference (EDOC '02) , 2002. 293-304.

[5] M Marchese. Service Oriented Architectures for Supporting Environments in eGovernment Applications[C] . 2003 Symposium on Applications and the Internet Workshops (SAINT '03 Workshops) , 2003. 106-110.

[6] Marine XML Schema Documentation[EB/OL] . http://www.aodc.gov.au/products/prod/documentation/marine_xml_schema.html.

[7] BODC Parameter Dictionary[EB/OL] . http://www.bodc.ac.uk/documents/bodc_params_download.html.

[8] Report of the Second Session of ICES-IOC Study Group on the Development of Marine Data Exchange Systems Using XML[EB/OL] . <http://www.marinexml.net/>.

[9] Report of the Third Session of ICES-IOC Study Group on the Development of Marine Data Exchange Systems Using XML[EB/OL] . <http://www.marinexml.net/>.

作者简介:

郭忠文(1965-) , 男, 教授, 博士生导师, 主要研究方向为计算机网络、海洋信息分布式处理技术; 尚传进(1980-) , 男, 硕士, 主要研究方向为移动 Agent 理论、传感器网络; 管恩花(1979-) , 女, 硕士, 主要研究方向为传感器网络、海洋信息分布式处理技术。

(上接第 150 页)

2.4 实验测试结果

分别在不同网络速度进行多次实验测试, 对不同等级的用户传送不同等级的图像, 取其中部分测试结果, 如表 5 所示。

表 5 不同等级用户传输时间

速度等级	网络速度(Bps)	传输时间(ms)
等级 0	2 152	657
	2 355	610
	2 865	484
等级 1	3 801	603
	4 265	594
	4 992	531
等级 2	5 364	719
	5 550	672
	6 044	656
等级 3	9 528	656
	10 300	609
	14 910	469
等级 4	19 920	625
	17 591	656
	21 750	516
	755 600(局域网)	15

3 结论

“金字塔”式网络传输模型根据用户等级不同而发送不同的数据量, 解决了各用户由于网速不同或网络速度的不稳定性而造成获取图像不同步的问题, 减少网络速度慢的用户的获取

图像等待时间。即使将来 JPEG 2000 完全代替了 JPEG, 该模型仍然可以继续使用。该模型灵活方便, 对于医疗图像、电子商务等相关领域中也有着参考价值, 特别是在多用户的图像网络传输, 多用户协同工作中有着重要的意义。

参考文献:

[1] hapiro J M. Embedded Image Coding Using Zerotrees of Wavelet Coefficients[J] . IEEE Transactions on Signal Processing, 1993, 41(12) : 3445-3462.

[2] K F Chen, *et al.* Analysis and Architecture Design of EBCOT in JPEG2000[J] . Proc . IEEE Int. Symp. on Circuits and Systems, 2001, (2) : 765-768.

[3] Tung-Shou Chen, Chen-Yi Lin. A New Improvement of JPEG Progressive Image Transmission Using Weight Table of Quantized DCT Coefficient Bits [C] . PCM 2002, LNCS 2532, 2002. 720-728.

[4] 宋长安, 王思贤, 吴柱, 等. Internet 上图像渐进传输的研究[J] . 计算机应用研究, 2001, 18(12) : 41-43.

[5] 张益贞, 刘滔. Visual C++ 实现 MPEG/JPEG 编解码技术[M] . 北京: 人民邮电出版社, 2002.

作者简介:

霍正聃(1980-) , 男, 硕士研究生, 主要研究方向为计算机网络、数字图像处理; 田地(1958-) , 男, 教授, 博士生导师, 主要研究方向为计算机在测控系统中的应用技术、分析仪器; 江游(1979-) , 男, 硕士研究生, 主要研究方向为图像数字图像处理网络化仪器; 熊行创(1979-) , 男, 工程师, 主要研究方向为计算机网络、计算机控制; 张玉海(1965-) , 男, 高级工程师, 主要研究方向为仪器科学与技术。