

基于多源异构数据的北京 GISC 数据请求服务

姚燕,李湘,郭萍,郑波

(国家气象信息中心,北京 100081)

摘要:为实现 WMO 信息系统(WIS)中全球信息系统中心(GISC)数据发现、数据获取以及数据订阅等主体功能中的数据请求服务,文章在分析多源异构气象数据源特点的基础上,针对不同的异构数据集成模式从实现和性能方面进行比较,采用 XML 和 Web Service 技术结合中间件模式对北京 GISC 中多源异构的气象数据进行集成访问、处理和请求服务的实现,并着重阐述了数据请求服务中各个功能层的详细设计内容和处理流程。数据请求服务在北京 GISC 业务系统底层气象数据源升级改造中发挥出了其低成本、可扩展和灵活性高的特点,使得多个数据业务系统、多种异构数据源并存,满足了北京 GISC 用户对气象数据的需求。随着 2013 年北京 GISC 业务系统建设验收和业务化运行,数据请求服务运行稳定,在北京 GISC 系统应用软件建设中取得了一定的成效。

关键词:数据集成;异构数据;中间件;XML;Web Service

中图分类号:TP311

文献标识码:A

文章编号:1673-629X(2021)08-0150-06

doi:10.3969/j.issn.1673-629X.2021.08.026

Data Request Service in Beijing GISC System Based on Multi-source Heterogeneous Meteorological Data

YAO Yan, LI Xiang, GUO Ping, ZHENG Bo

(National Meteorological Information Center, Beijing 100081, China)

Abstract: In order to realize the data request service in the data discovery, data retrieval and data subscription, which are the main functions of GISC (global information system center) in the WIS (WMO information system), the characteristics of multi-source heterogeneous meteorological data are analyzed firstly, then the implementation and performance comparison of different heterogeneous data integration model are shown. On the basis of those, the Beijing GISC data request service is implemented by XML and Web Service technology combined with middleware model of multi-source heterogeneous meteorological data. Furthermore, the detailed design and processing flow in various functional layers are emphatically expounded. In the upgrading and improvement of the underlying meteorological data sources in Beijing GISC system, the technical advantages of data request service, such as low cost, scalability, high flexibility and so on, have been shown, enabling multiple data service operational systems and heterogeneous data sources to coexist and meeting the demand of Beijing GISC users for meteorological data. With the operation of Beijing GISC operational system in 2013, the data request service is stable, and certain achievements have been made in the construction of application software of Beijing GISC system.

Key words: data integration; heterogeneous data; middleware; XML; Web Service

0 引言

二十一世纪初,世界气象组织(world meteorological organization, WMO)各项计划已经拥有或正在开发各自独立的气象信息系统。系统的多样性不仅使系统之间缺乏兼容和效率,而且存在重复建设和高成本等弊端。为了避免这些问题的进一步加剧, WMO 提出了 WMO 信息系统(WMO information system, WIS)的概念,旨在建立一个通用的、综合的信

息服务平台,用以支撑各项计划以及相关组织的气象数据交换和共享^[1]。全球信息系统中心(global information system center, GISC)是 WIS 的核心功能中心,其主要功能之一就是要提供与数据存储位置无关的气象数据发现、访问和检索服务^[2]。

国家气象信息中心从 2009 年开始逐步开展 WIS 实施,在 2011 年完成北京 GISC 系统主体功能的建设,并通过 WMO 的能力评估和测试,成为首批全球信息

收稿日期:2020-10-04

修回日期:2021-02-05

基金项目:科技部 2009 年度公益性行业(气象)科研专项(GYHY200906042)

作者简介:姚燕(1979-),女,硕士,高级工程师,研究方向为气象数据处理与分析。

系统中心之一。随着 WIS 实施的推进,北京 GISC 收集和提供服务的数据种类和数量也将进一步拓展和增加^[3]。但是由于业务需求和功能归属的不同,体系内许多不同气象数据内容的信息服务和管理系统已经逐步建立起来。这些系统的数据源有着各自不同的处理对象、操作方法和专用客户端,基本上相互隔离。因此迫切需要在多源异构的数据环境下设计一个统一的数据集成服务来支撑北京 GISC 用户对气象数据的访问和检索,实现北京 GISC 业务系统中各类气象数据的请求功能^[4]。

1 服务设计

1.1 数据分析

在北京 GISC 系统中,数据请求服务所涉及到的气象资料主要涉及全球交换的常规气象资料、全球交换的二进制气象资料(包括 GRIB、BUFR 等不同数据格式^[5])、中国本地的全球数值预报模式产品、中国风云卫星资料产品以及交互式全球大集合数值预报产品(THORPEX Interactive Grand Global Ensemble^[6], TIGGE)等。通过分析归纳起来整个服务数据源的异构性主要体现在以下三个方面:

(1)系统异构,即数据源所依赖的业务应用系统、数据库管理系统之间的不同构成了系统异构。北京 GISC 系统的数据源涉及到业务系统实时气象数据库、本地系统应用数据库以及 TIGGE 数据专用存储检索系统 MARS(meteorological archival and retrieval system)。

(2)接口异构,即数据源在检索接口上存在不同。实时气象数据库系统需要采用其统一定制的检索服务 API(application programming interface,应用程序接口);TIGGE 数据检索则需要安装和使用其专用的客户端软件;本地数据库则可以通过 JDBC 等接口直接进行数据表的 SQL 检索。

(3)来源异构,即内部数据源和外部数据源之间的异构。

北京 GISC 系统要实现这些数据的访问和检索服务,就必须为数据请求提供一个公共的统一的集成接口,使访问者不必关心如何才能获取到,也不必考虑数据的来源和异构性、数据抽取、数据合成等问题,只需指定请求的数据条件即可。

1.2 技术分析

目前最主要的异构数据集成体系结构有以下三种^[7]:联邦数据库模式、数据仓库模式和中间件模式。

联邦数据库模式支持分布、异构与自治,其对应系统包括一组互相协作但分别自治的数据库系统,这些自治的数据库系统称为成员数据库系统。这些成员数

据库系统可以进行不同程度的集成,而且数据依然保留在原来的存储位置,不必构建一个集中式的数据仓库。随着数据源的增多,数据查询效率会受到限制,主要适用于异构数据源较少的数据集成^[8]。

数据仓库模式主要是建立一个存储数据的仓库,由数据抽取、转换与装载工具定期地从数据源过滤数据,然后装载到数据仓库中,提供用户操作。所有的查询都针对数据仓库中的数据,数据仓库必须根据数据的变化而随时更新。但数据仓库中数据实时性不强,该模型主要适用于对数据联机分析和决策支持系统的集成^[9]。

中间件模式为用户提供了一个全局模式,用户提交的查询都是针对这个全局模式而进行的,因此数据源的位置、模式及访问对用户来说都是透明的。系统将基于中间模式的查询转换为针对各局部数据源的查询,将用户的查询分解为各个数据源的子查询,并将各子查询返回的数据综合起来得到查询结果。其优点是数据不需冗余,而且能保证是最新的。中间件模式适合于数据变化较多的异构数据源集成,局部数据源的加入操作实现比较灵活方便,而且可以保持各个数据源充分的自治性。

1.3 总体结构

综合以上数据分析和数据集成模式分析,考虑中间件模式成本低、易实现以及局部数据源加入方便灵活、自治性强等优点,北京 GISC 系统采用 Web Service 和 XML 技术并应用中间件数据集成模型来进行多源异构气象数据请求服务的设计和实现。数据请求服务的总体结构如图 1 所示,总体从逻辑上分为三层,分别是应用层、应用服务层和数据源层^[10]。

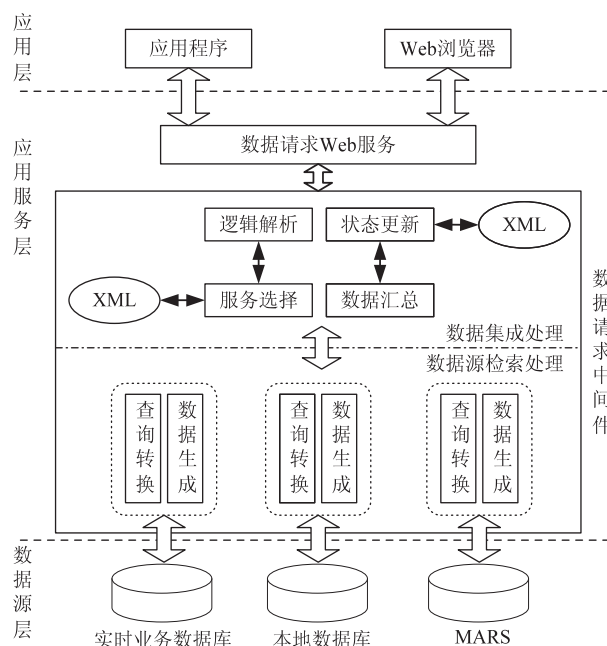


图1 数据请求服务总体结构

应用层是终端用户与系统交互的接口。该层根据具体的应用和用户环境,采用合适的信息访问技术或应用软件,可以是 Web 浏览器或专用的客户端,来提供用户相应的数据请求信息,并从应用服务层接收相应的请求结果状态信息和结果数据文件。无论应用是 C/S 模式还是 B/S 模式,只要遵循应用服务层的数据请求接口规范,就可以有效地、透明地从底层的数据源获取数据。

应用服务层负责完成所有数据请求逻辑处理,也是总体结构中的核心层,主要包括 Web 服务和数据请求中间件^[11]。数据请求 Web 服务收集来自应用层的数据请求信息,并为每个数据请求生成全局唯一的任

务编号。它首先检查该数据请求是否在数据缓存中,即同样的数据是否已被处理过。如果存在则直接反馈用户结果信息和数据。否则根据用户请求数据内容不同生成请求条件描述接口文件,转发给数据请求中间件。数据请求中间件进而解析请求条件描述接口文件,完成服务选择、数据检索、数据压缩或分拆、合并结果、状态更新等一系列处理后,将请求结果状态信息和结果数据返回给 Web 服务。应用层通过任务号可以查询到请求任务的执行情况,并获得相应的请求执行结果。相应的处理流程如图 2 所示。

数据源层主要包括需要访问的各个自治的局部数据源,为数据请求服务提供各种气象数据支撑。

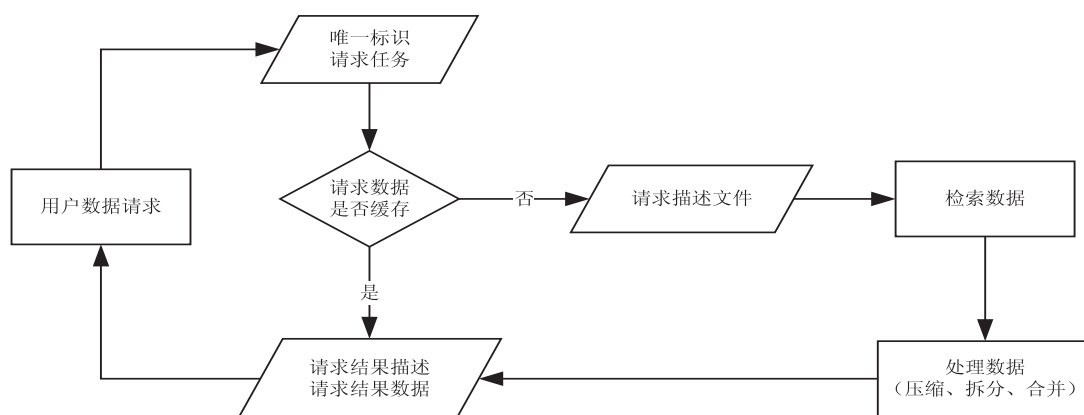


图 2 数据请求处理流程

2 关键技术

2.1 Web Service

Web Service 是基于一组标准 Internet 协议的分布式计算组件,具有开放、面向 Internet 标准化接口等特点,能够实现松散耦合的、与平台无关的应用系统交互与协同,是应用集成的理想平台。它使用 SOAP (simple object access protocol) 表示信息传输协议,使用 WSDL (web service description language) 进行本身内容描述,使用 UDDI (universal description, discovery and integration) 来发现、描述与集成 Web Service^[12]。Web Service 为数据集成提供了灵活的访问方式,能够较好地数据集成提供标准的开发接口和良好的扩展性,为快速新增和部署新数据源提供了方便。

2.2 XML 技术

可扩展标记语言 (extensible markup language, XML) 作为一种结构性标记语言,描述能力的扩展性强,结构化程度高,具有标准的编程接口,因其简单、可扩展、跨平台的特性而被广泛应用于数据信息的交换和传输^[13]。由于 XML 具有良好的互操作性及数据表示能力,它可以表示任何形式的数据^[14],所以在北京 GISC 数据请求服务中均采用 XML Schema 来定义数据请求服务相关的接口参数信息及请求结果信息。请求参数信息主要涉及请求数据集名称、各个数据集对应请求条件以及请求结果存放目录和文件名等信息,如表 1 所示。而请求结果信息则主要包括请求结果状态信息及请求结果文件,如表 2 所示。

表 1 数据请求条件信息 XML 内容

XML Schema	定义内容	数据说明/示例
<dataset>	请求数据集标识	global exchange 代表全球资料 CMA NWP 代表数值预报产品 CMA NSMC 代表卫星资料 TIGGE 代表 TIGGE 数据
<condition>	请求参数条件信息	可以包含多个 AHLinfo 或 TGLinfo 信息
<AHLinfo>	请求资料的简式报头信息	主要描述非 TIGGE 数据信息

续表 1

XML Schema	定义内容	数据说明/示例
<startdate>	请求数据开始时间	YYYYMMDD(年月日)
<enddate>	请求数据结束时间	YYYYMMDD(年月日)
<datatimes>	请求资料时次信息	HHMM(时分)
<TT>	请求数据公报报头信息	例如“SM”
<AA>	请求数据公报报头信息	例如“CI”
<ii>	请求数据公报报头信息	例如“01”
<CCCC>	请求数据来源	例如“BABJ”
<TGLInfo>	请求 TIGGE 数据信息	主要描述 TIGGE 数据信息
<CCCC>	TIGGE 资料中心	例如“LFPW”
<date>	TIGGE 资料日期	可设置多个值,以“,”分隔
<datatimes>	TIGGE 资料时次	可设置多个值,以“,”分隔
<number>	TIGGE 资料集合数	可设置多个值,以“,”分隔
<ltype>	TIGGE 资料层次	可设置多个值,以“,”分隔
<param>	TIGGE 资料要素	可设置多个值,以“,”分隔
<area>	TIGGE 资料地理范围	可进行区域裁剪,例如“china”
<resultdir>	结果数据存放位置	以唯一任务号命名的工作目录
<maxsize>	用户请求数据量的最大阈值	从用户信息获取,单位 MB

表 2 数据请求结果信息 XML 内容

XML Schema	定义内容	数据说明/示例
<status>	请求结果状态标识	<0:请求处理出错
		0:请求已提交,正在进行处理;
		1:请求处理成功,数据已获取;
		2:请求数据超大;
<resultfile>	请求结果文件	3:请求数据不存在
		请求结果文件名或文件名列表

2.3 数据请求中间件

数据请求中间件由数据集成处理和数据源检索处理两部分构成。它是一个组合服务,能够解析接收到的 XML 语言描述的数据检索描述文件,完成一系列服务后,将检索结果信息和数据文件统一地返回给 Web 服务器和应用层。因此数据请求中间件技术是实现北京 GISC 系统数据请求服务的关键^[15]。

数据集成检索代理集成了四个核心服务,包括逻辑解析、服务选择、结果汇总和状态更新。服务解析服务主要是负责对接收到的检索描述 XML 文件进行分析,为检索服务的选择提炼出相应的接口信息或关键字。服务选择服务负责根据所提炼的关键字对服务元数据进行检索,找到对应的提供检索服务的数据源,然后将服务请求向相应的数据源检索接口服务转发。结果汇总服务负责将数据源检索接口服务返回的结果数据和处理信息进行汇总。状态更新服务实现根据汇总后的处理信息和结果更新检索请求的元数据信息,这样 Web 服务器能够实时地将最新的处理状态和结果

返回给用户。

数据源检索接口代理主要进行针对异构数据源的检索访问操作。每一个不同的异构数据源都有一个不同的检索服务与之相对应,通过对不同的异构数据源的服务化封装,起到屏蔽异构数据源的目的,并将封装的服务提供上层数据集成检索代理的服务调用。

每个异构数据源对应的检索服务均包括查询转换和数据生成两部分功能。不同数据源支持的查询语言是不同的,因此需要依次提取出其中的查询参数,转换为针对各个局部数据源能够识别的查询语句或脚本。例如:实时数据库的查询需要将查询条件和内容整理为其统一检索接口所规定的输入参数列表形式。MARS 的查询需要将查询条件和内容整理为其所能处理的检索脚本格式。而本地数据则需要将检索条件组织成相应关系型数据库的标准的 SQL 查询语句即可。数据生成服务则调用相关数据源 JDBC 或处理 API,来执行各个局部数据源的数据检索,获取相应的数据,统一以数据文件的形式返回给上层服务。

3 应用效果

采用 XML 和 Web 服务结合中间件架构的设计,初步实现了北京 GISC 系统中异构数据源的透明集成检索和请求服务,使得多种业务应用系统、多种异构数据源并存,并随着北京 GISC 业务系统建成已业务化运行多年,为用户提供了稳定的数据请求服务。在北京 GISC 系统中,数据请求服务主要应用于数据发现、

数据获取和数据订阅这三个主要功能中。

在数据发现功能中,注册用户检索到所需要的气象数据元数据信息时,用户可以直接点击数据访问按钮直接浏览到对应的气象数据文件,同时也能通过“是否获取更多数据”来调用数据请求服务从后台数据源中获取用户指定时间的所查元数据对应数据文件,如图 3 所示。

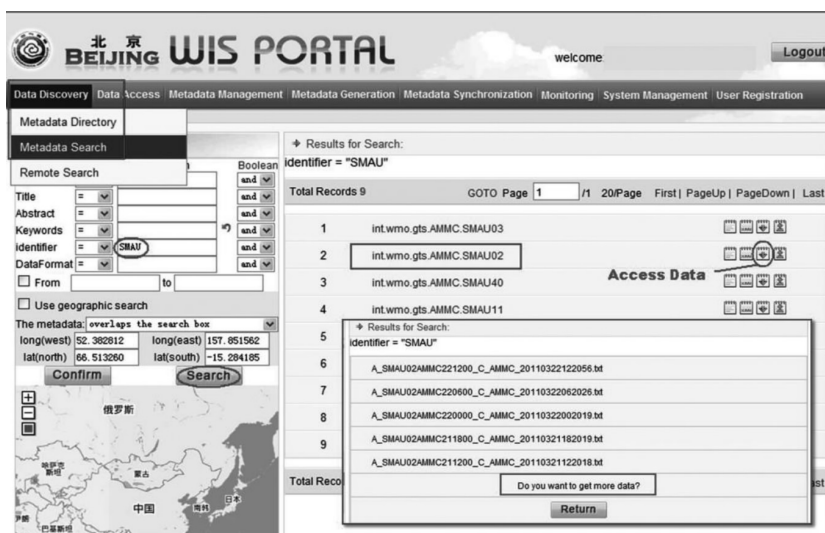


图 3 数据请求服务在数据发现功能中的应用示例

在数据获取功能中,注册用户可以根据需求在数据目录中选择不同的气象数据种类,然后根据不同数据种类中对应的请求要素条件进行设置,最终通过数

据请求服务实现气象数据文件压缩包的下载,如图 4 所示。

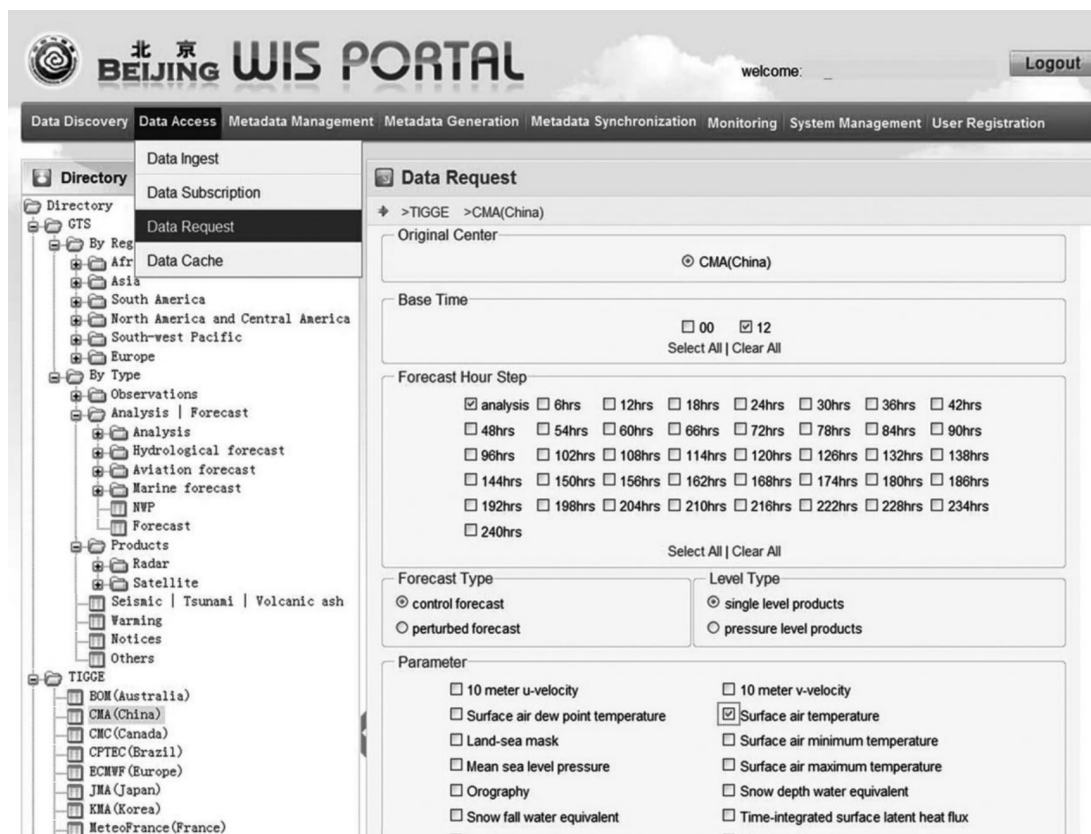


图 4 数据请求服务在数据获取功能中的应用示例

在数据订阅功能中,注册用户通过选择所需要的数据内容,加入数据订阅购物车,提交数据订单。数据订阅后台应用根据资料时间特性,通过定时器依据资料类别触发数据请求服务,获取订阅结果气象数据,并以用户指定方式返回数据订阅用户,如图 5 所示。

Check	Identifier	Time Group
<input type="checkbox"/>	cn.gov.cma.NMC.NWP.BABJ.HCXD85	00
<input type="checkbox"/>	cn.gov.cma.NMC.NWP.BABJ.HCXD99	00
<input type="checkbox"/>	cn.gov.cma.NMC.NWP.BABJ.HCXD70	00
<input type="checkbox"/>	cn.gov.cma.NMC.NWP.BABJ.HCXD20	00
<input type="checkbox"/>	cn.gov.cma.NMC.NWP.BABJ.HCXD50	00
<input type="checkbox"/>	cn.gov.cma.NMC.NWP.BABJ.HCXD92	00

图 5 数据请求服务在数据订阅功能中的应用示例

4 结束语

长期以来,由于异构数据源之间的显著差异,使得异构气象数据集成请求成为一个难点问题。中间件技术的使用,为解决异构数据集成请求服务提供了机会。同时这种服务设计也保护了原有信息化投资,能够适应底层数据源应用系统由旧向新的平滑过渡,比如实时业务数据库在近几年的更新换代,GISC 系统中 48 小时 Cache 数据源的加入,能够满足系统建设的低成本、阶段性和扩展性需求,在北京 GISC 系统第一版应用软件建设中取得了一定的成效。同时随着近几年 WIS2.0 概念的提出以及关于气象云数据的共享服务的逐步实施^[16],也为北京 GISC 业务系统第二版的建设提供了技术经验。

参考文献:

- [1] World Meteorological Organization. WIS project & implementation plan (v1.2) [EB/OL]. 2010. https://www.wmo.int/pages/prog/www/WIS/ref_docs_en.html.
- [2] World Meteorological Organization. Manual on the WMO information system: annex VII to the WMO technical regulations (2019 edition) [EB/OL]. 2019. <http://library.wmo.int/>.
- [3] 李 湘,王甫棣,姜立鹏,等. WIS 的实现技术研究及应用[J]. 气象,2011,37(10):1301-1308.
- [4] World Meteorological Organization. WMO information system functional architecture (v1.0) [EB/OL]. 2019. https://www.wmo.int/pages/prog/www/WIS/ref_docs_en.html.
- [5] WMO SECRETARIAT. Manual on codes-international codes, volume I. 2, annex II to the WMO technical regulations: part B - binary codes, part c - common features to binary and alphanumeric codes (2019 Edition) [EB/OL]. 2020. <http://library.wmo.int/>.
- [6] THORPEX 中国委员会. 观测系统研究与可预报性试验中国计划(THORPEX-China) [M]. 北京:气象出版社,2006.
- [7] 姚崇东. 基于 XML 的多源异构数据集成的实现方法研究[D]. 哈尔滨:哈尔滨工程大学,2007.
- [8] 邓劲生,郑倩冰. 信息系统集成技术[M]. 北京:清华大学出版社,2012.
- [9] 苏 琪,刘西林,王 军. 基于 Web Service 的数据集成研究及应用[J]. 计算机技术与发展,2014,24(8):51-54.
- [10] 杨 辰,周立新,毕奔腾. 基于 SOA 的岩溶地质数据服务系统设计与实现[J]. 计算机技术与发展,2019,29(4):149-153.
- [11] 吴兴华. 基于 SOA 架构的铁路旅服系统集成平台服务中间件设计与实现[J]. 铁路计算机应用,2018,27(3):1-4.
- [12] 黄序鑫. 基于 XML 和 Web Service 的数据集成研究与应用[J]. 现代计算机:专业版,2012(6):39-41.
- [13] 冯少奇,巩思亮,陈春明. 一种基于 XML 的海量视频监控信息检索技术[J]. 计算机技术与发展,2016,26(4):182-186.
- [14] 吴家菊,唐定勇,席传裕. 基于 XML 的信息集成研究[J]. 计算机工程与设计,2007,28(20):5018-5020.
- [15] 赖 欣,胡敬玉. 基于 AIXM 的民航共享数据集成系统构建[J]. 计算机技术与发展,2020,30(6):181-185.
- [16] World Meteorological Organization. WMO information system 2.0 strategy (2017 Edition) [EB/OL]. 2018. <http://library.wmo.int/>.