

何佳,裴顺强,惠建忠,等.基于RESTful架构的气象数据检验接口设计与实现[J].气象研究与应用,2024,45(1):49-55.
HE Jia, PEI Shunqiang, HUI Jianzhong, et al. Design and implementation of meteorological data inspection interface based on RESTful architecture[J]. Journal of Meteorological Research and Application, 2024, 45(1): 49-55.

基于RESTful架构的气象数据检验接口设计与实现

何佳,裴顺强,惠建忠*,梅钰,丰德恩,高金兵,郝江波

(中国气象局公共气象服务中心,北京 100081)

摘要:气象数据质量评价是整个数据链条中的关键一环。为解决目前气象数据检验业务分散滞后、资源不集约及服务效率不高的问题,实现统一服务,基于现代Web应用流行框架RESTful,首次将检验作为Web服务,基于共性检验需求设计并实现RESTful架构风格的气象数据检验接口。统一接口服务将业务资源进行集中,提高物理资源利用率的同时释放人力资源。接口服务以及基于接口开发的检验系统,可实现检验结果秒级在线查询,大大提高检验效率,丰富的接口和可视化系统给用户带来良好体验。

关键词: RESTful; OAS; connexion; 时序数据库; 检验评估

中图分类号: P409

文献标识码: A

doi: 10.19849/j.cnki.CN45-1356/P.2024.1.09

检验评估是提升气象服务产品质量的重要抓手。借助实时检验评估结果,动态掌握产品质量,为自身产品优化改进及基于检验结果进行优势产品遴选的动态模型构建提供客观依据。目前,检验评估算法研究较为成熟,中国气象局在2005^[1]、2009^[2]和2012^[3]年发文对检验算法有明确描述,许多算法在业务中广泛应用^[4-9]。为方便业务人员开展检验,国家气象中心研发算法库并向社会开放^[10]。但有关检验评估系统平台的研究相对滞后。组织内部以产品团队为单位开展检验,资源不集约,效率低下,无法满足实时获取结果的需求。2019年中国气象局公共气象服务中心成立检验团队,开展检验平台技术研究。采用“自定义格式+内存数据库+高并发Web框架方案”开发检验数据处理子系统。由于自定义格式通用性差,扩展性不高,且方案涵盖技术路线多,应用成本大。检验算法虽然是公开明确的,但在实际应用时有很多需要确定的细节。比如,如何定义有效样本;多分类检验又可分为多分类和分级检验;样本筛选时间是按

观测数据时间还是预报起报时间等。细节在具体实施时最终明确,这就可能存在各家采用不同方案,检验结果不能横向比较,造成应用的局限性。综上,气象服务数据检验存在“效率低、滞后严重、应用受限”的问题。

公共气象服务以分众化聚焦服务对象,已开展面向公众、新能源、交通、生态等领域的气象服务。不同领域气象服务最终产品差异较大,而其背后依赖的基础数据产品是共用的。本文以基础数据产品为检验对象,以公众、交通、能源气象服务产品为例,分析梳理检验需求如下:

(1)产品分类检验。不同行业服务需求决定了多样的产品类型。如交通气象服务包括实况、短临预报和短期预报,能源气象服务有数值预报、场站订正预报、功率预测等。

(2)气象要素检验。每类产品包括一至多个气象要素,不同领域关注要素不同。如交通服务关注气温、能见度、降水、相对湿度等,能源服务关注辐射、风速等,公众服务关注天气现象、最高或最低气温等。

收稿日期: 2023-12-21

基金项目: 中国气象局能力提升联合研究专项(23NLTSZ009)

第一作者简介: 何佳(1983—),女,高级工程师,主要从事气象信息技术研究和项目管理工作。E-mail: hejia_1002@163.com

*通讯作者: 惠建忠,正高级工程师,从事气象信息系统设计与研发工作。E-mail: 1404713480@qq.com

(3)检验评估方法。每个气象要素可以采用一至多个评估指标,不同产品的相同要素检验方法可以不同。

(4)阈值分级配置。交通行业气温在 -4°C 以下、 $-4\sim-2^{\circ}\text{C}$ 情况下,道路可能结冰而影响交通安全,业务更关注此范围的气温预报能力。能源行业功率预测准确率需结合装机容量、切入切出风速进行计算,不同场站风机规格不同参数不统一。

(5)预报时效质量。目前交通服务中短临预报产品的时效分辨率为10 min,短期预报则为1 h,能源服务产品为15 min,公众服务实况为1 h,预报为12 h,不同产品时效分辨率不同。其次,用户端有其自身关注的时效范围,如风电场用户关注第1 d和第3 d时效数据质量。

(6)时空范围定义。时间和空间是气象数据的基础属性,也是检验样本筛选的基本条件。时间条件由产品时间和分辨率确定,空间条件由业务具体确定,如公路交通以道路为空间范围,能源气象以发电厂位置为空间范围。

基于上述共性需求,从资源集约、业务效率、横向比较等角度切入,创建高效、统一的检验服务是本文实现的目标。

1 关键技术

基于RESTful架构的气象数据检验接口总体技术框架如图1所示。采用前后端分离开发方式,前端负责Web应用可视化交互界面开发,后端负责接口开发。检验算法使用国内开源气象检验算法库MetEva+自研算法相结合。为提高开发效率及规范性,采用“规范优先”的Python Web API框架connexion。使用时序数据库实现数据存储管理。

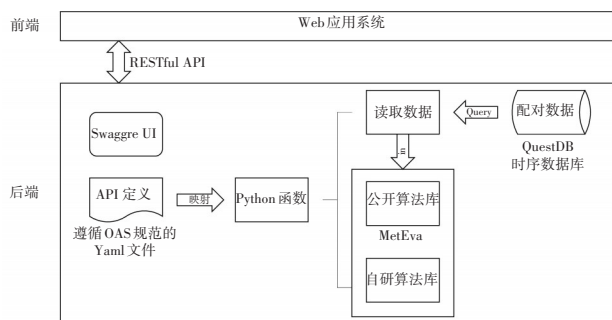


图1 基于RESTful架构的气象数据检验接口总体技术框架

1.1 RESTful软件架构

REST(Representational State Transfer, 又称表述

性状态移交^[11])是由Roy Fielding在其2000年的博士论文中提出。RESTful软件架构非常适合前后端分离开发方式中作为连接前后端的接口开发技术。许皓皓等^[12]采用微服务框架设计和开发标准化气象数据服务接口,通过RESTful API统一发布数据,改变应用程序直接读取数据库的旧模式,降低应用系统开发成本。

本文首次将气象数据检验作为一种Web资源,以Web服务方式将气象数据检验服务化。用户通过HTTP协议访问接口获取检验结果,访问过程仅需明确检验接口参数值,省去了过去准备检验数据、编写检验算法、搭建运行环境等重复工作。

1.2 Python Web框架

Connexion是一个基于Open API规范(简称OAS)自动处理http请求的框架。Connexion创新性的使用“从规范到代码”,提高API开发速度的同时,保障代码与文档的一致性。使用Connexion作为Web开发框架可以大大简化开发过程,快速构建符合工业标准的业务API。实际开发过程可概括为两个步骤:第一步定义API和实现端点函数;第二步编写端点函数。Connexion集成Swagger,用于描述和文档化RESTful API,接口部署后通过Servers/UI地址进行访问。可视化交互界面提供接口名称、接口参数、参数属性等信息,接口使用者输入参数值点击“执行”即可看到结果。

1.3 气象数据检验算法

国家气象中心研发的“天气预报全流程检验评估程序库(Meteorological Evaluation Program Library, 简称MetEva),集成5类54种检验算法,涵盖世界气象组织推荐和国内业务规范要求的大部分检验算法^[10],是目前国内权威性、全面性较高的一款检验工具。何佳等^[13]基于MetEva库设计气象服务产品检验流程,实现定期输出检验报告。检验接口开发依该库作为检验算法工具。重点应用基础层中的检验算法库中分类预报检验、连续型预报检验,功能层的数据预处理模块和检验分析模块中的数值型检验指标。当业务上出现新的检验方法时,在代码层面借助Python模块机制将新算法以独立模块进行组织,解决与已有算法冲突。在架构层面,新算法的数据结构、函数参数等参照MetEva进行设计,保持总体上检验算法的一致性。

1.4 数据存储管理技术

时间序列数据是按数据属性中的时间戳大小

排列的一系列记录值的数据,常见的时序数据如金融、气象、交通等行业数据。基于时序数据库特点,在数据写入性能、查询性能和存储压缩率等指标上对其进行针对性优化,相比传统关系型数据库在性能上有很大提升。近年来基于时序数据库的应用研究越来越多。郝建明等^[14]基于时序数据库建立高速公路数据集成平台,解决高速公路海量时序数据高效压缩存储和快速实时分析难题。倪昱等^[15]基于时序数据库建立的海洋装备监控数据存储系统,在存储空间和查询性能上相较传统关系数据优

势显著。气象数据是典型的时序数据,采用时序数据库作为检验服务的数据管理技术,可在读写性能方面发挥更为出色的表现。以交通气象服务基础数据检验为例,初步统计数据量见表1。每月数据达 2×10^8 条,使用传统关系数据库难以满足实时检验需求。QuestDB是2014年开源的一个面向列的关系型时序数据库,本文基于QuestDB存储配对数据(时间对齐的标准数据和待检验数据对),通过查询接口为检验算法提供输入。

表1 交通气象服务基础数据检验数据量分析

产品类型	产品要素	时间分辨率	空间分辨率
实况分析		每 10 min	公里级公路桩号点(1 km 分辨率)
短临预报	能见度、风速、风向、	未来 40 min, 逐 10 min	小时能见度 2 617 点, 小时其它 1 726 点
短期预报	降水、气温、相对湿度	未来 72 h, 逐小时	分钟能见度 1 762 点, 分钟其它 1 098 点
数据量汇总		小时能见度: 1.35×10^8 条/月, 小时其它: 8.9×10^7 条/月 $\times 5$ 要素 分钟能见度: 3.8×10^7 条/月, 分钟其它 2.4×10^7 条/月 $\times 5$ 要素	

每类产品每个要素设计为一张表。表命名规则为:行业领域+产品类别_气象要素,例如:公路交通短期预报气温数据库表名为 roadocf1h_tem。每张表前 4 列为坐标信息依次为 level、time、dtime、id,分别为数据样本的层次、时间、预报时效、站号。从第 5 列开始为数据部分。数据部分第一列为观测值(图 2 中 obs 列),后面为预报值。如果同一类产品来自多个不同系统则对应多个预报值,由多个数据列进行存储(图 2 中 hfjqx、nmc、qhmsc 分别代表三家成员),基于列的存储设计技术为对比检验提供了可行性。

坐标信息部分				数据部分			
层次	起报时间	预报时效	站号				
level	time	dtime	id	obs	hfjqx	nmc	qhmsc
0	2023-08-19 00:00:00	0	113943824	3.6	3.583 0	3.5	3.689 6
0	2023-08-19 00:00:00	15	113943824	4.59	4.022 5	3.6	3.744 6
0	2023-08-19 00:00:00	30	113943824	4.45	4.471 6	3.7	3.884 8
0	2023-08-19 00:00:00	45	113943824	3.77	4.928 2	3.8	4.015 2

图2 检验匹配数据表结构

(图中数据仅为辅助说明,非真实数据)

2 接口设计与实现

2.1 接口组成

接口设计的一个重要原则是可扩展性,可扩展

的背后是一系列标准规范。目前已经出台的《QX/T 618-2021 气象数据服务接口规范》^[16]和《气象业务软件组件化开发指南(2023版)》^[17]对接口组成进行了规范。

(1) 接口名称

清晰的接口名称能够更好的表明接口目的。“检验”表达核心功能,检验对象(数据产品、气象要素)、检验算法(亦称检验评价指标/方法)、时空范围等是核心功能的细化。将描述功能的要素进行提炼、组合,实现接口命名清晰表达接口功能的目的。

检验接口命名由核心功能、检验方法、分组维度和主要条件等 4 部分,并给出各部分取值规则(见表 2)。

(2) 接口参数

接口中的参数本质上是一种变量技术,表示意义相同但取值不同的对象。参数设计合理则可以减少接口数量。如在检验业务中,当某类产品某个要素有多个检验指标时,则可以将检验指标作为接口参数。实际调用时使用一个接口,通过给参数赋不同值实现一个接口即可获得多种检验方法结果。接口参数在 URL 定义中用一对花括号表示:如/{product}/tem/{method}/,示例中 tem 是固定的,表示气温要素,product 和 method 是路径参数,标识产品和检验方法。

表2 检验接口名称各部分取值规则

组成部分	代码	含义	备注
核心功能	check	气象数据检验	本文描述的所有接口使用统一代码。
检验方法	指标缩写	检验评价指标	1、使用指标英文缩写。 2、如果接口仅返回一个指标数据,则使用代码列表中的一种。 3、如果接口返回全部指标数据,则可使用“method”标识。 4、示例:TS、MAE、RMSE、CORR 等
分组维度	GroupXXX	检验样本分组维度	1、使用固定单词Group+“维度”进行标识。 2、维度参考 meteva 在线文档。 3、示例: 1)按预报时效分组 GroupDtime 2)按天分组 GroupDay
主要条件	{ByXXX}And{ByXXX}	主要参数条件	1、使用固定单词“By”和条件名称进行标识。 2、示例: 1)按时间范围筛选样本 ByTimeRange 2)按时效范围筛选样本 ByDtimeRange

(3)返回码

接口返回码的定义由两部分组成:返回码取值和含义。返回码分为两大类,成功和失败,编码由业务逻辑定义。遵守“应详细描述失败原因,便于接口用户理解后端实际发生的情况”的原则,检验接口返回码设计见表3。

(4)返回数据结构

接口返回数据采用 json 格式,数据结构统一封装,最高层级包括三部分:检验结果(“checkResult”标签)、返回码(“messageCode”标签)、返回码描述(“messageInfo”标签)。检验结果内部根据场景灵活设计。如多家产品对比检验,标签代表生产系统;多分类检验,标签表示类别;分组检验,标签表示分组。多方式组合检验,如多产品、多方法检验,使用嵌套技术外层表示检验方法,内层表示参与检验的产品。

2.2 业务接口设计

满足共性需求的前提下,考虑行业需求差异,

表3 检验接口返回码

返回码类别	返回码取值	返回码含义
成功	0	成功
	1	无数据
	2	观测缺测
失败	3	数据异常(给出异常码:如-998)
	4	数据超出阈值

检验接口按照服务行业进行分类设计。

(1)交通服务气象数据检验接口

交通气象数据有三类产品,每类产品检验要素相同,每个要素检验方法有交叉,每类产品每个要素检验方法相同。接口设计按要素分类,即每个要素设计一个接口。产品类型、检验方法、时空范围等作为接口参数。

以能见度要素为例,交通气象数据检验接口 URL 设计如下:

http://IP:PORT/v1/vis/{product}/{method}/接口名

表4 交通气象数据检验需求

序号	气象要素	所属产品	检验方法
1	能见度	实况分析 短临预报 短期预报	分级TS评分、分级偏大/小率、绝对误差分级TS评分
2	风速		分级TS评分、分级偏大/小率
3	风向		分级TS评分
4	降水量		晴雨TS评分、分级TS评分、分级偏大/小率
5	温度		分级TS评分、分级偏大/小率、绝对误差分级TS评分
6	相对湿度		绝对误差分级TS评分

称?query_params

(2)能源服务气象数据检验接口

能源气象数据检验包括三类产品,每类产品包

括的要素不同,每个要素的检验方法不完全相同。

接口设计按产品分类,即每类产品设计一个接口

(表5)。

表5 能源气象数据检验需求

序号	产品类别	气象要素	检验方法
1	模式预报	辐射、风速	相关系数、平均误差、平均绝对误差、均方根误差
2	模式订正	辐射、风速	相关系数、平均误差、平均绝对误差、均方根误差、准确率、合格率
3	功率预测	功率	相关系数、平均误差、平均绝对误差、均方根误差、准确率、合格率、95%分位数偏差率
4	所有类别	——	到报率、及时率

以模式预报为例,能源气象数据检验接口 URL

设计如下:

仓库,服务端拉取镜像、运行。逻辑部署图见图3。

数据内容检验接口 : http://IP:PORT/v1/model/{area}/{element}/接口名称?query_params

数据文件评价接口 : http://IP:PORT/v1/接口名称?query_params

(3)公众服务气象数据检验接口

公众服务以 APP 发布的实况和预报两类数据

作为评价对象。评价要素包括气温和天气现象。

实况类取整点数据,预报类仅评价 08 和 20 两个时

次。考虑预报类接口查询参数多一个时次,设计时

将两类产品分开(表6)。

表6 公众气象数据检验需求

序号	产品类别	气象要素	检验方法
1	实况	整点气温	均方根误差
2		整点天气现象	命中率
3	预报	前/后 12 h 气温	均方根误差
4		天气现象	命中率
5	所有类别	——	到报率、及时率

公众气象数据检验接口 URL 设计如下:

实况产品检验接口 : http://IP:PORT/v1/obs/{element}/{method}/接口名称?query_params

预报产品检验接口 : http://IP:PORT/v1/forecast/{element}/{method}/接口名称?query_params

3 检验接口测试与应用

3.1 接口测试

接口响应时间是衡量接口性能的重要指标,是提高业务效率的直接反映。首先完成接口部署,采用 Docker 方式将接口程序打包成镜像,上传到私有

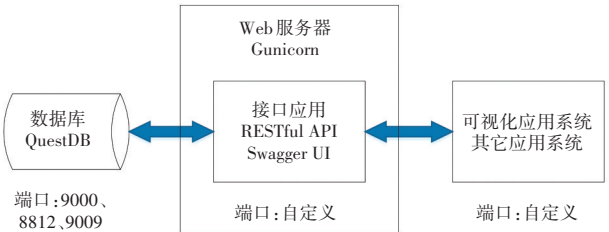


图3 检验接口逻辑部署图

接口部署后分别从交通、能源、公众三个行业每个行业选取一个接口进行测试。时间范围选取日、周、月三个维度,每个接口随机调用5次,计算平均耗时,结果见表7—9。从结果可以看出:(1)所有接口响应耗时均在秒级;(2)除交通短临产品接口外,其它接口均到达毫秒级;(3)接口响应时间随样本数量(等价时间范围长度)呈现增大趋势;(4)除样本数量因素外,接口响应时间受指标数量、算法复杂度、是否分组等多因素影响。

3.2 接口应用

接口的使用者可以分为两类。一类是开发者用户,是接口的直接访问者。第二类用户是最终用户,通过访问开发者基于接口开发的各类应用系统(前端),是接口的间接使用者。前端根据用户需求进行 UI 设计,以丰富的交互和展现方式提供检验结果,例如提供时间、空间作为基本查询条件,配合产品、要素、检验方法进行组合。展示方式包括 GIS 地图、折线图、柱状图、表格等,如使用 GIS 地图展示站点检验结果在空间上的分布,使用折线图展示多家产品按时效检验结果的对比情况,通过表格展示分级检验各级别结果等。

开发前端系统属于项目级别,还可以实现更轻

表7 单日检验接口响应时间

单位:s

行业	接口	样本数	第一次	第二次	第三次	第四次	第五次	平均
能源	模式, 4个指标, 96个时效总体	97	0.032	0.029	0.032	0.029	0.033	0.031
公众	7 d 预报, 1个指标, 单时效	2 150	0.072	0.086	0.074	0.069	0.066	0.073
交通	短临, 3个指标, 4个时效独立	5 760	0.251	0.233	0.248	0.250	0.241	0.245

表8 单周检验接口响应时间

单位:s

行业	接口	样本数	第一次	第二次	第三次	第四次	第五次	平均
能源	模式, 4个指标, 96个时效总体	679	0.064	0.047	0.045	0.052	0.048	0.051
公众	7 d 预报, 1个指标, 单时效	14 939	0.157	0.165	0.164	0.162	0.153	0.160
交通	短临, 3个指标, 4个时效独立	24 000	1.596	1.677	1.612	1.601	1.691	1.635

表9 单月检验接口响应时间

单位:s

行业	接口	样本数	第一次	第二次	第三次	第四次	第五次	平均
能源	模式, 4个指标, 96个时效总体	2 894	0.119	0.094	0.094	0.090	0.092	0.098
公众	7 d 预报, 1个指标, 1个时效	64 564	0.563	0.634	0.548	0.523	0.562	0.566
交通	短临, 3个指标, 分时效检验	208 878	13.34	12.91	12.71	12.64	12.52	12.824

量的应用。如针对统计一段时间内逐日检验结果排名前三的需求,首先调用接口获取每天结果进行排名,然后统计排名天数。在统计数据基础上绘制柱状图(图4)。

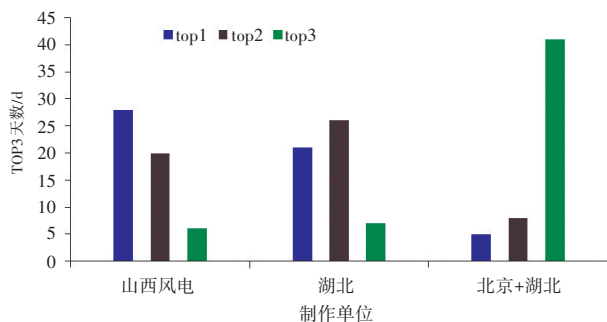


图4 检验接口二次开发应用排名统计柱状图
(2023年9月1日至10月31日短时临近功率山西风电场
准确率TOP3天数)

4 结论与讨论

基于RESTful架构开发的气象数据检验接口首次将检验服务化、集中化,实现物理资源有效集约。集中化服务的提供,替代了各产品团队独立检验,解放了人力资源。高效的接口性能支持高分辨率

产品(分钟级)长时间序列检验结果的快速获取,提高业务效率的同时带给用户良好体验。“多产品对比检验”的接口设计实现“用一把尺子进行测量”,打破各自独立检验结果不能横向比较的限制,对比结果可在商务服务中发挥重要作用。目前接口已开放给内部人员使用。下一步将从安全性角度为接口访问设置权限管理,便于扩大使用范围。

参考文献:

- [1] 中国气象局. 中短期天气预报质量检验办法[Z]. 2005.
- [2] 中国气象局预报与网络司. 全国城镇天气预报质量国家级检验方案[Z]. 2009.
- [3] 中国气象局预报与网络司. 中短期天气预报质量检验工作改革方案[Z]. 2012.
- [4] 刘丽敏, 石磊, 赵波, 等. 中央台精细化温度预报在伊春市的检验和误差分析[J]. 黑龙江气象, 2018, 35(2): 24-25.
- [5] 万夫敬, 赵传湖, 马艳, 等. ECMWF模式气温预报在青岛地区的检验与评估[J]. 气象科技, 2018, 46(1): 112-120.
- [6] 邹阳, 王将. 2个数值模式在昆明地区气温预报中的准确率比较[J]. 安徽农业科学, 2018, 46(10): 140-144.
- [7] 包慧濛, 郭达烽, 肖潇. ECMWF集合预报统计量在江西降水预报中的检验分析[J]. 气象与减灾研究, 2015,

- 38(3):60-67.
- [8] 罗布坚参, 普布顿珠, 除多, 等. 三种数值预报模式降水产品在青藏高原检验分析[J]. 高原山地气象研究, 2015, 35(2):24-28.
- [9] 郑淋淋, 邱学兴. 一种改进的降水临近外推预报技术方法研究及效果检验[J]. 气象科技, 2020, 48(1):97-106.
- [10] 刘凑华, 代刊, 林建, 等. 天气预报全流程检验评估程序库的设计与实现[J]. 气象, 2022, 48(7):813-827.
- [11] 理查德森, 阿蒙森. RESTful Web APIs 中文版[M]. 赵震一, 李哲, 译. 北京: 电子工业出版社, 2014:404.
- [12] 许皓皓, 姚日升, 沃伟峰. 标准化气象数据服务接口设计与实现[J]. 气象科技, 2018, 46(4):685-691.
- [13] 何佳, 惠建忠, 袁亚男, 等. 基于 MetEva 的气象服务产品检验流程设计与实现[J]. 气象研究与应用, 2022, 43(2):127-132.
- [14] 郝建明, 袁逸涛. 基于时序数据库的高速公路数据集成平台[J]. 上海船舶运输科学研究所学报, 2022, 45(4):57-63.
- [15] 倪昱, 郑志恒. 基于时序数据库的海洋装备监控数据存储系统[J]. 现代信息科技, 2023, 7(11):101-104.
- [16] 中国气象局. 气象数据服务接口规范: QX/T 618-2021[S]. 北京: 中国气象局, 2021.
- [17] 中国气象局预报与网络司. 气象业务软件组件化开发指南(2023 版)[EB/OL]. (2023-09-13)[2023-11-27]. https://www.cma.gov.cn/zfxxgk/gknr/wjgk/qtwj/202309/t20230913_5772613.html.

Design and implementation of meteorological data inspection interface based on RESTful architecture

HE Jia, PEI Shunqiang, HUI Jianzhong*, MEI Yu, FENG Deen, GAO Jinbing, HAO Jiangbo
(Public Meteorological Service Centre, China Meteorological Administration, Beijing 100081, China)

Abstract: The quality evaluation of meteorological data is a key link in the entire data chain. In order to solve the status quo of dispersion, lagging, inefficient resources and low service efficiency in meteorological data inspection business, and to achieve unified services, inspection is used as a Web service for the first time based on the popular framework of modern Web applications RESTful, and it is designed and implemented into a RESTful architecture style of meteorological data inspection interface based on common inspection requirements. The unified interface service centralizes business resources, and improves the utilization rate of physical resources while releasing human resources. Interface services and the interface-based inspection system achieve the second-level online query of the inspection results, greatly improving the efficiency of the inspection, and the rich interfaces and visualization system bring users a good experience.

Keywords: RESTful; OAS; Connexion; time Series database; testing and evaluation