

王刚,俞春强.提高智能网络中地震数据集成与共享性能的研究[J].地震工程学报,2020,42(2):416-421.doi:10.3969/j.issn.1000-0844.2020.02.416

WANG Gang, YU Chunqiang. Improvement of Integration and Sharing Performance of Seismic Data in Intelligent Network[J]. China Earthquake Engineering Journal, 2020, 42(2): 416-421. doi: 10.3969/j.issn.1000-0844.2020.02.416

# 提高智能网络中地震数据集成与共享性能的研究

王 刚<sup>1</sup>, 俞春强<sup>2</sup>

(1. 桂林旅游学院 网络信息中心, 广西 桂林 541006;

2. 广西师范大学 网络信息中心, 广西 桂林 541004)

**摘要:** 地震数据集成共享在地震监测与分析中具有重要作用。为此,针对传统基于数字图书馆的地震数据集成化共享方法集成度低、共享率差的问题,研究一种新的地震数据集成与共享方法,该方法分为两个阶段,先是利用数据仓库技术对地震数据进行集成,包括数据抽取、清洗、归约、装载等;后建立共享平台,完成数据透明式访问,实现数据共享。结果表明:利用该文章所提出的方法对大规模地震数据进行集成化共享,集成度达到 95.68%,共享率达到 87.34%,为遥测地震监测与预防提供了有效的数据查询平台。

**关键词:** 智能化网络; 遥测地震数据; 集成; 共享方法

中图分类号: P315.61

文献标志码: A

文章编号: 1000-0844(2020)02-0416-06

DOI: 10.3969/j.issn.1000-0844.2020.02.416

## Improvement of Integration and Sharing Performance of Seismic Data in Intelligent Network

WANG Gang<sup>1</sup>, YU Chunqiang<sup>2</sup>

(1. Network Information Center, Guilin Tourism University, Guilin 541006, Guangxi, China;

2. Network Information Center, Guangxi Normal University, Guilin 541004, Guangxi, China)

**Abstract:** Seismic data integration and sharing play a significant role in seismic monitoring and analysis. The conventional digital library-based seismic data integration and sharing method have the problems of low integration and poor sharing rates, so a new seismic data integration and sharing method are studied. The method is divided into two stages. First, seismic data is collected using data warehouse technology, including data extraction, cleaning, reduction, loading, etc.; then a sharing platform is set up to complete the data transparent access, and the data sharing is realized. The results show that the proposed method's degree of integration and sharing rate can reach 95.68% and 87.34%, respectively, providing an efficient data query mechanism for seismic monitoring and prevention of telemetry.

**Keywords:** intelligent network; telemetry seismic data; integration; sharing method

收稿日期: 2019-07-09

基金项目: 国家自然科学基金项目(61762017)

第一作者简介: 王刚(1968-),男,湖南新宁人,在职研究生,高级实验师,研究方向: 计算机应用。E-mail: wg@gltu.edu.cn。

通信作者: 俞春强(1988-),男,硕士研究生,助理研究员,研究方向: 图像处理。

### 0 引言

我国位于世界两大地震带——环太平洋地震带与欧亚地震带之间,受太平洋板块、印度板块和菲律宾滨海板块的挤压,地震断裂带十分活跃,尤其对于一些处于地震带上的山地地区来说,地质活动频发,给人们带来了严重的生命财产损失,因此每年国家在地震遥感监测上都花费大量的时间和金钱,以减少灾害带来的损失。在这种情况下,产生了大量的各类地震数据,如震害数据、地震区划数据、地质数据、地震台网数据、历史地震数据、地震目录数据等<sup>[1]</sup>。这些数据对于地震规律分析、灾害预防预警有着重要作用。然而,由于这些数据存在不同的业务系统当中,数据储存格式和管理方式都大不相同,导致无法实现数据集成共享,影响了地震防御决策的制定与实施。为此,研究遥测地震数据进行集成化共享方法具有重要的实际意义。

目前,存在的地震数据集成化共享方法多是利用数字图书馆技术完成异构数据集成化共享,但是这种传统的集成共享方法只适用小规模的数据信息,一旦数据规模过大,其方法的集成度和共享率就会降低<sup>[2]</sup>。

针对上述传统方法存在的问题,研究一种新的地震数据集成化共享方法。该方法主要分为两个阶段:第一阶段利用数据仓库进行地震数据集成,包括抽取数据、处理数据等过程;第二阶段进行遥测数据共享,主要包括服务注册、共享创建以及查询转换等。最后为测试方法的有效性进行仿真测试。结果表明:与传统地震数据集成化共享方法相比,本方法的集成度和共享率有了提高,为地震监测与预防提供了重要的技术支持<sup>[3]</sup>。

### 1 地震数据集成化共享方法

地震数据的采集主要是通过遥测的手段。遥测是将对象参量的近距离测量值传输至远距离的测量站来实现远距离测量的技术,能获取中国大陆大范围 and 时空密集的地壳运动观测数据。其特点如下:第一监测范围广泛;第二监测反应迅速,能实时反映目标的动态变化;第三监测项目种类丰富,可以获得不同的目标特征;第四不受地形等条件的限制,获取信息方便。遥测主要应用于气象、土地、海洋、农业、地质和军事等领域<sup>[4]</sup>。

目前为更好地监测地震,我国设有地震监测预报、震灾防治和紧急救援三大工作体系,因此监测出来的地震数据存在不同的业务系统当中,导致数据

储存格式和管理方式都大不相同,从而造成了虽然累积大量的地震数据资源,但是由于缺乏有效的管理手段,存在大量的数据散失和共享壁垒的问题,数据共享的程度比较低。此外,异构地震数据的存在,使得大量地震数据难以归类存储,相关信息的查准率和查全率有待提高<sup>[5]</sup>。为此研究采集到的地震数据集成化共享方法具有重要意义。图 1 为本次研究的智能化网络中地震数据集成化共享方法基本思路。

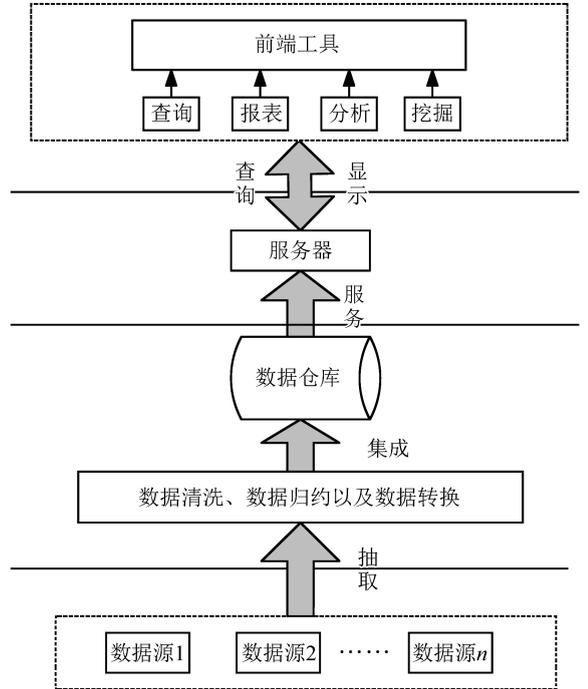


图 1 基于数据仓库的地震数据集成化共享基本思路  
Fig.1 Basic framework of seismic data integration and sharing based on data warehouse

#### 1.1 地震数据集成

数据集成是指将不同来源、格式、特点性质,但又互相关联的分布式异构数据集中到一起的过程,以使用户能够以透明的方式访问这些数据源,从而达到数据共享的目的。

数据集成的难点与重点在于数据异构性、分布性以及自治性上如表 1 所列。

表 1 数据集成特性

Table 1 Data integration characteristics

特性	说明
异构性	不同业务系统当中的数据都是独立的,因此在数据语义、表达形式以及使用环境上等均不相同。
分布性	数据异构往往伴随着数据异地分布,所以需要通过网络进行传输。
自治性	各个数据源有很强的自治性,它们可以在不停止集成系统的前提下改变自身的结构和数据,给数据集成系统的鲁棒性提出挑战。

以上三个遥测地震数据的特性是阻碍集成的核心问题,为此相关专家研究了很多解决措施,如联邦式数据库、中间件集成方法、数据仓库方法等。其中,联邦式数据库是最早期人们采用的一种集成方法,优点是集成化程度较高,用户参与少;缺点是构建一个全局数据模式的算法复杂,扩展性差。中间件集成方法是一种更为先进的数据集成方式,优点是应用范围广泛,不仅对于结构化数据有用,对集成半结构化数据同样有用,且有很好的查询性能,自治性强;缺点是无法像联邦式数据库那样支持双向读写功能,它通常是只读的<sup>[6]</sup>。针对以上两种集成方法存在的问题,在本研究中选用第三种数据仓库集成方法,它是利用复制原理将多源数据集中到一起,然后用户就可以像访问普通数据库那样直接查询数据,实现资源共享。其基本流程如图2所示。

从图2中可以看出,数据仓库集成技术基本思路:从不同数据源中抽取原始地震数据到基础数据层当中,同时其进行清洗、转换、装载等,将不同格式、形式、来源数据转化为统一格式的地震数据,并对其进行装载,迁移储存到数据仓库当中。

### (1) 数据抽取

数据抽取是数据进入数据仓库的入口,主要包括互联、复制、增量、转换、调度等环节,目前对于数据抽取主要通过抽取工具来完成,其原理是通

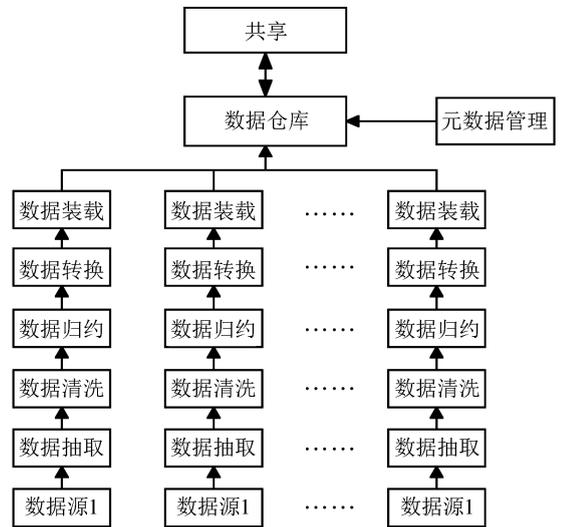


图2 数据仓库集成数据流程

Fig.2 Data integration flow of data warehouse

过建立源数据与目标数据之间的对应关系生成抽取代码,利用计算机强大的计算功能进行自动抽取,同时对抽取到的数据进行清洗、转换等<sup>[7]</sup>。在这里主要由 ETL 工具完成。

目前国内外市场上主要 ETL 工具有 Ascentia 公司的 DataStage、Oracle 公司的 OWB、微软公司的 DTS、IBM 公司的 DWC 等,这些工具的性能比较如表2所列。

表2 四种 ETL 工具性能比较

Table 2 Performance comparison of four ETL tools

来源	ETL 工具	优点	缺点
Ascentia 公司	DataStage	支持内部编程和工作流	提取功能相对简单,效率和精度有待提高
Oracle 公司	OWB	提供模型设计和构建以及元数据管理	使用过程复杂,缺乏灵活性
Microsoft 公司	DTS	支持工作流和多个数据源	不能跨平台
IBM 公司	DWC	支持集群提取,尤其是数据量越大,效率越高	数据源越复杂,就越耗时;用户界面一般

从数据源中进行数据抽取的同时,由于有的数据历时时间长,数据失去效用,以及还存在许多错误、冗余、重复等数据,因此需要对数据进行处理,包括数据的清洗、归约以及转换。

### (2) 数据清洗

数据清洗是指发现并纠正数据文件中可识别的错误,以增强数据集成质量,其具体流程如图3所示。

### (3) 数据归约

面对海量的地震数据,缩减其规模对于降低存储空间占有率具有重要作用,可以减少共享平台负载,提高运行效率。数据归约是指从目标大数据中提取具有代表性的数据特征,以此代替原始大数据,

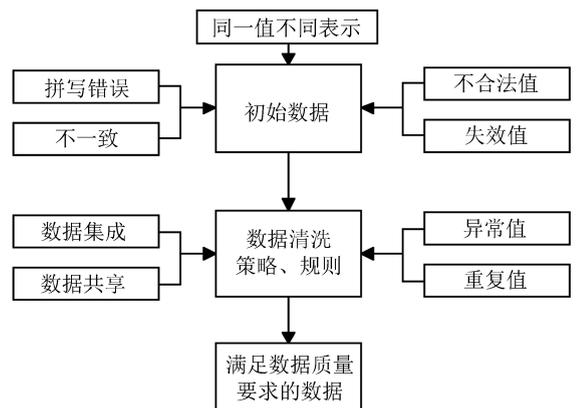


图3 数据清洗具体流程

Fig.3 Data cleaning process

从而在尽可能保持数据原貌的前提下,最大限度地

精简数据量,降低数据规模<sup>[8]</sup>。目前,数据归约主要有三种方法:特征归约、样本归约与特征值归约。在这里由于面对不相关、冗余、重复等大规模已知地震数据,因此选择样本归约方法降低地震数据规模。样本归约就是从数据集中选出一个有代表性的样本的子集,特点是成本少、速度快、范围广,有时甚至能获得较高的精度。

(4) 数据转换

由于集成到的数据多源异构,并不能直接统一储存到数据仓库当中,无法实现共享<sup>[9]</sup>,因此需要将其转换成适合数据共享形式,主要包括平滑、聚集、泛化和规范化处理(图 4)。

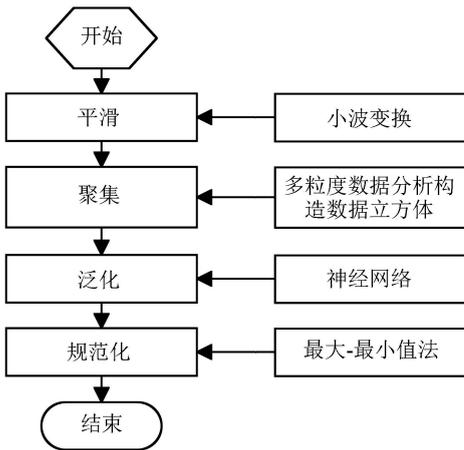


图 4 数据转换基本思路

Fig.4 Basic framework of data conversion

(5) 数据装载

以上环节处理地震数据结束后,需要将其加载到数据仓库当中,以便为后续共享做准备。数据加载是指运行接口程序和数据仓库传输规则,将数据装入到数据仓库当中,主要过程如图 5 所示。

1.2 遥测地震数据共享

遥测地震数据集成结束后,接下来进行数据共享。数据共享是让在不同地方使用不同计算机、不同软件的用户能够读取他人数据并进行各种操作运算和分析,在这里主要通过构建数据共享平台来实现。在该平台中,主要包括三个功能模块:服务注册模块、共享创建模块以及查询转换模块(图 6)<sup>[10]</sup>。

(1) 服务注册模块

服务注册模块主要作用是为用户提供注册服务,然后创建用户权限,最后用户可以在自己权限范围内查询地震资源。具体包括在线用户管理、用户注册和映射规则维护等三部分,如图 7 所示<sup>[11]</sup>。

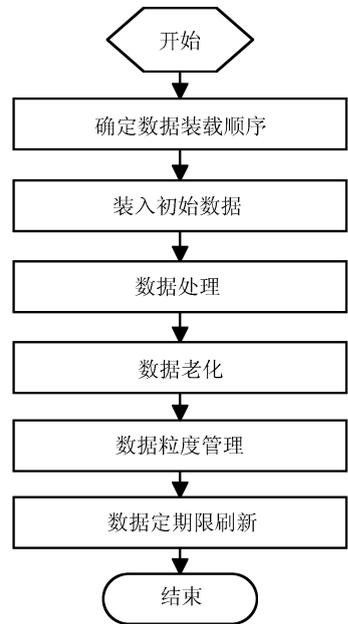


图 5 数据装载基本流程

Fig.5 Basic flow of data loading

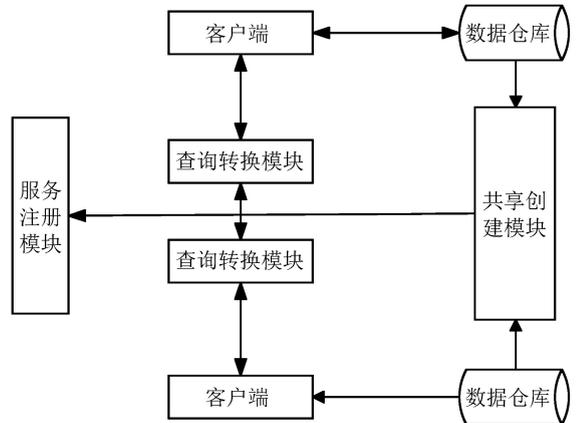


图 6 遥测地震数据共享

Fig.6 Telemetry seismic data sharing

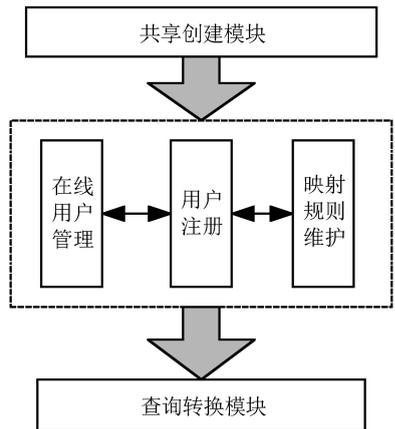


图 7 服务注册模块

Fig.7 Service registration module

## (2) 共享创建模块

共享创建模块主要作用是创建用户指令与数据仓库之间映射程序,即如何根据用户指令,识别数据仓库中与之相关的数据,并将其呈现给用户<sup>[12]</sup>。该模块主要包括如图8所示几部分。

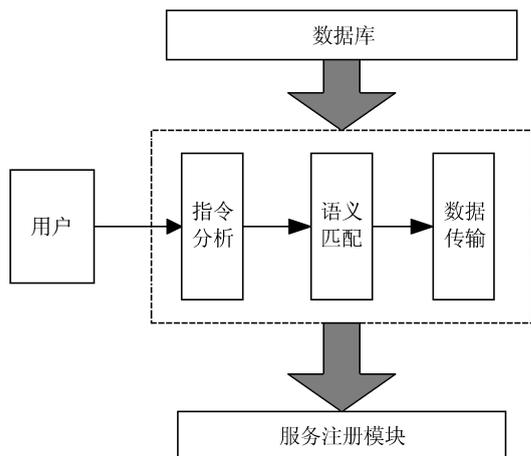


图8 共享创建模块

Fig.8 Share creation module

## (3) 查询转换模块

查询转换模块主要作用是为用户查询提供服务,即将根据指令识别出的数据通过统一接口传输到共享窗口当中。具体组成如图9所示。

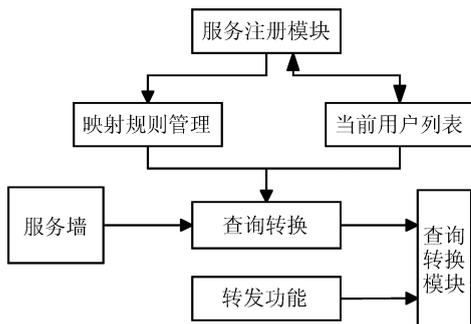


图9 查询转换模块

Fig.9 Query conversion module

## 2 方法性能测试

性能仿真测试是验证方法是否达到需求目标的关键步骤。在这里根据传统基于数字图书馆遥测地震数据集成化共享方法存在的问题:集成度和共享率低,对本次研究的基于数据仓库的遥测地震数据集成化共享方法进行性能测试。

数据集成度,即多源异构地震数据集成后,有效数据占总数据的百分比,数值越高,集成度越高。

数据共享率,即同一时间共享平台允许正常访

问的最多人数占总人数(平台无法运行时的人数)的比例,数值越大,方法共享性能越好。

本次测试地震数据来自国家地震科学数据共享中心(<http://data.earthquake.cn/>)。利用B/S结构,即Browser/Server(浏览器/服务器)结构将地震数据导入模拟软件,模拟来自547个数据源,数据种类有187种,总数据规模为145TB;访问人数以50为基数,然后以每次增加10人的速度进行共享访问,直至平台无法承载为止;数据集成工具选择Oracle公司提供的OWB和广州卓一信息科技有限公司提供的云数据图书馆;数据共享平台采用Microsoft Visual Basic 7.0 行程开发,如图10所示。



图10 数据共享平台

Fig.10 Data sharing platform

现在利用两种方法进行集成共享测试,结果如表2所列。

表2 方法集成度与共享率测试结果

Table 2 Test results of integration and sharing rate of methods

测试方法	集成度/%	共享率/%
传统方法	87.34	81.25
本文方法	95.68	87.41

由表2可知,对来自547个数据源的145TB遥测地震数据进行集成和共享,利用本文方法的集成度达到95.68%,共享率为87.34%;而利用传统方法的集成度为87.41%,共享率为81.25%。通过以上两种方法的结果对比可知,本文方法的集成与共享性能更好。

## 3 结束语

综上所述,地震作为一种无法规避的自然灾害,

我们能做到的只是尽量降低灾害带来的损失,因此每年我国在地震灾害监测方面付出了极大的努力,产生了大量的监测数据。由于这些数据来自不同的业务部门,彼此相互独立,共享利用率低,不利于地震分析。为此,本次在传统地震数据集成化共享方法基础上,研究一种基于数据仓库的遥测地震数据集成化共享方法,该方法将数据仓库技术用于其中,提高了数据的集成度;建立共享平台,分模块设计数据共享体系,实现数据透明式共享,为地震灾害监测与预防的业内研究人员提供参考。

## 参考文献(References)

- [1] 刘坚,马文娟,李盛乐,等.基于大数据技术的地震科学数据集成共享研究[J].中国科技资源导刊,2017,49(5):74-81.  
LIU JIAN, MA WENJUAN, LI SHENGLE, et al. Integration and Sharing Research on the Earthquake Science Data Based on the Technology of Big Data[J]. China Science & Technology Resources Review, 2017, 49(5): 74-81.
- [2] 郑钰,彭朝勇,杨建思.历史地震图数据库及共享平台[J].地震地磁观测与研究,2017,38(4):207-211.  
ZHENG Yu, PENG Chaoyong, YANG Jiansi. Historical Seismogram Database and Sharing Platform Based on the Distributed File System[J]. Seismological and Geomagnetic Observation and Research, 2017, 38(4): 207-211.
- [3] 董一兵,王想,张环曦,等.基于测震强震仪器健康状态数据的应用集成开发[J].地震工程学报,2017,39(1):177-185.  
DONG Yibing, WANG Xiang, ZHANG Huanxi, et al. Integrated Application Development Based on Health State Data from Seismic & Strong-motion Instruments[J]. China Earthquake Engineering Journal, 2017, 39(1): 177-185.
- [4] 冯玉苹,徐维秀,刁瑞,等.面向地震勘探算法的模块集成技术研究[J].地球物理学进展,2017,32(5):2214-2221.  
FENG Yuping, XU Weixiu, DIAO Rui, et al. Method of Module Integration for Seismic Exploration Algorithm[J]. Progress in Geophysics, 2017, 32(5): 2214-2221.
- [5] 姚庆华,和永军,缪应锋.面向综合智能交通系统的多源异构数据集成框架研究[J].云南大学学报(自然科学版),2017,39(S1):41-45.  
YAO YAO Qinghua, HE Yongjun, MIAO Yingfeng Study on the Integration Framework of Multi-source Heterogeneous Transportation Data[J]. Journal of Yunnan University(Natural Sciences Edition), 2017, 39(S1): 41-45.
- [6] 杨超,胡尧,商明菊,等.基于SV模型的行程时间预测[J].贵州大学学报(自然科学版),2019,36(5):11-17.  
YANG Chao, HU Yao, SHANG Mingju, et al. Prediction of Travel Time Based on SV Model[J]. Journal of Guizhou University(Natural Sciences), 2019, 36(5): 11-17.
- [7] 鲜学丰,崔志明,方立刚,等.面向Deep Web本地化数据集成的数据源两层选择模型[J].计算机工程,2017,43(3):32-39.  
XIAN Xuefeng, CUI Zhiming, FANG Ligang, et al. Data Source Two-layer Selection Model for Deep Web Localized Data Integration[J]. Computer Engineering, 2017, 43(3): 32-39.
- [8] 刘坤,陈幼佳,章敏.地震条件下桥台主动土压力的合理简化计算方法[J].太原理工大学学报,2016,47(2):249-253,258.  
LIU Kun, CHEN Y. Frank, ZHANG Min. Simplified Approach for Rational Determination of Seismic Induced Earth Pressure on Abutments[J]. Journal of Taiyuan University of Technology, 2016, 47(2): 249-253, 258.
- [9] 薛腾飞,傅群超,王枫,等.基于区块链的医疗数据共享模型研究[J].自动化学报,2017,43(9):1555-1562.  
XUE Tengfei, FU Qunchao, WANG Cong, et al. A Medical Data Sharing Model Via Blockchain[J]. Acta Automatica Sinica, 2017, 43(9): 1555-1562.
- [10] 邓仲华,黄雅婷.“互联网+”环境下我国科学数据共享平台发展研究[J].情报理论与实践,2017,40(2):128-132.  
DENG Zhonghua, HUANG Yating. Research on the Development of Scientific Data Sharing Platform at Home in the “Internet +” Environment[J]. Information Studies (Theory & Application), 2017, 40(2): 128-132.
- [11] 谢有顺,李盛乐,刘小利,等.面向地震行业的地震数据共享服务平台的设计与实现[J].大地测量与地球动力学,2017,37(5):546-550.  
XIE Youshun, LI Shengle, LIU Xiaoli, et al. Design and Implementation of Seismological Data Sharing Service Platform for Seismological Industry User[J]. Journal of Geodesy and Geodynamics, 2017, 37(5): 546-550.
- [12] 李勤,裴兴旺,孟海.某砖混厂房火灾后安全性评定与再生设计[J].消防科学与技术,2019,38(7):1034-1036.  
LI Qin, PEI Xingwang, MENG Hai. Safety Assessment and Regeneration Design of a Brick-concrete Factory after Fire[J]. Fire Science and Technology, 2019, 38(7): 1034-1036.