



大数据时代地质信息化建设及技术探讨

李敏¹, 傅洁², 李磊¹, 王小丹¹, 陈致远¹, 邓凡¹,
黄垒¹, 曾乐¹, 王国明¹, 王心华¹

1. 中国地质调查局 天津地质调查中心, 天津 300170; 2. 中国地震局第一监测中心, 天津 300180

摘要: 基于多年信息化建设经验, 结合信息化技术的发展与地质工作的实际需求, 在概述国内外地质调查信息化发展现状的基础上, 重点阐述了地质资料信息化服务历程、地质信息产品研制、地质大数据存储管理、地学知识图谱、地质大数据可视化、区块链的地学应用场景等, 以期借助大数据等技术更好地管理数据资产、推动地学数据共享、完善地质信息化建设。

关键词: 地质信息; 大数据; 地质时空数据; 知识图谱; 可视化; 区块链

Geological informatization construction and technology in the big data era

LI Min¹, FU Jie², LI Lei¹, WANG Xiao-dan¹, CHEN Zhi-yuan¹, DENG Fan¹,
HUANG Lei¹, ZENG Le¹, WANG Guo-ming¹, WANG Xin-hua¹

1. Tianjin Center of China Geological Survey, Tianjin 300170, China;

2. The First Monitoring and Application Center, CEA, Tianjin 300180, China

Abstract: Based on years of experience in informatization construction, combined with the development of information technology and practical demands of geological work, this paper summarizes the current development status of informatization in geological surveys at home and abroad, and focuses on the evolution of geological data information services, development of digital geological information product, geological big data storage and management, geoscience knowledge graphs, geological big data visualization, and specific application scenarios of blockchain in geosciences, aiming to use big data technology to enhance data asset management, facilitate geoscience data sharing, and advance geological informatization construction.

Key words: geological information; big data; geological spatiotemporal data; knowledge graph; visualization; blockchain

随着新一代信息技术的飞速发展、产业格局的深刻变革、国家战略发展需要以及经济社会发展, 地质信

息化工作如箭在弦, 异常紧迫. 当前, 地质工作进入科学研究第四范式的新时代, 工作模式由粗放型、单一

收稿日期: 2023-11-16; 修回日期: 2024-01-03. 编辑: 周丽.

基金项目: 中国地质调查局项目“国家地质大数据汇聚与管理”(DD20190382).

作者简介: 李敏(1984—), 女, 硕士, 高级工程师, 主要从事地质信息化方面研究工作, 通信地址 天津市河东区大直沽八号路 4 号, E-mail//782768001@qq.com

通信作者: 傅洁(1986—), 女, 硕士, 工程师, 从事大数据理论、方法、技术研究工作, 通信地址 天津市河东区一号桥耐火路 7 号, E-mail//782768001@qq.com

引用格式: 李敏, 傅洁, 李磊, 王小丹, 陈致远, 邓凡, 黄垒, 曾乐, 王国明, 王心华. 大数据时代地质信息化建设及技术探讨[J]. 地质与资源, 2025, 34(4): 517-526.

Li M, Fu J, Li L, Wang X D, Chen Z Y, Deng F, Huang L, Zeng L, Wang G M, Wang X H. Geological informatization construction and technology in the big data era [J]. Geology and Resources, 2025, 34(4): 517-526.

型单要素调查向精准型、综合型多要素调查转变,需要对产生的结构化、非结构化、半结构化数据综合分析,建立数据—信息—知识—服务—再数据的大数据链,挖掘百年来积累的海量珍贵地质数据,推动数据共享,深化数据应用,服务经济社会发展,促进地质科普及地质文化传播。

1 国内外地质信息化发展概述

1.1 国际地质信息化进展

目前,国外各种信息技术已广泛应用于野外地质调查数据采集、属性数据管理、成果图件出版、数据一体化管理等方面。多数发达国家的地质信息综合应用已从单一的数据库建设、构建简单的应用系统过渡为分布式大型数据库技术、大型GIS技术、Web服务技术的综合应用,基本形成了地理信息共享与互操作的标准规范体系。美国联邦数据委员会提出以空间数据框架作为各种信息采集、配准、集成的通用基础。英国、加拿大、澳大利亚、新西兰、荷兰、马来西亚、日本、韩国等国家先后开始实施国家空间数据基础设施(National Spatial Data Infrastructure, NSDI)。欧盟议会和欧盟理事会以在欧盟层面及各参与国层面立法的方式颁布欧盟空间信息基础设施(INSPIRE),基于坚实的顶层设计框架,以履行法律义务形式引发欧盟、各参与国、各地学机构地理信息基础设施建设与地质信息共享服务现代化。它在本质上不是关于技术的基础设施,而是要在政府、企业、用户之间形成一个明确的框架协议,便于用户最大限度使用所需的公共信息、地理信息。

在数据共享方面,为了在网络上创建开放的、动态的全球地质图数据库,并使用GeoSciML标准共同使用地质数据,更好地服务社会,2008年8月6日,在挪威首都奥斯陆召开的第33届国际地质大会正式提出“OneGeology”计划,即“同一个地质计划”。OneGeology的开展促进了“玻璃地球”建设的全球合作。目前世界各国正在大力开展的“玻璃地球”是实现地质科学大数据一体化、集成化存储与管理的重要途径与基本方式。世界各国虽然都采取三维区域地质填图与深部探测计划相结合的方式建设“玻璃地球”,但进展程度、侧重面、方法、思路等各不相同。澳大利亚以发现下一代巨型矿床为目的开展了四维地质填图^[1];荷兰以建立全

荷兰地下地质学信息的数据银行为目的启动了地下数据和信息数字化填图项目,建立了全国1000m以浅的三维地层框架模型^[2];加拿大将三维地质填图用于盆地地下水调查^[3];英国通过建立多种不同分辨率的三维地层框架模型实现从概率模型到大比例尺精细模型的过渡^[4];法国在地质调查多领域针对不同用途利用不同的商业软件包开展了三维地质建模^[5];美国为探索北美大陆的构造与演化,揭示地震、火山喷发的秘密,开展为期15年的“地球透镜”计划,并针对资源环境评价开展三维地质框架研究^[6]。此外,还有瑞士、意大利地壳探测计划,德国大陆反射地震计划,俄罗斯深度地壳探测计划等。面对环境变化和资源短缺等巨大挑战,2016年,联合国教科文组织全球地球化学国家研究中心启动“化学地球”大科学计划,与世界各国一道共同实施地球五大圈层化学元素的含量和分布数据整理,持续记录全球化学元素基准与变化等科学数据,绘制地球化学元素图谱,将元素周期表上所有化学元素的含量和分布绘制在地球上,建立“化学地球”大数据平台,支撑全球自然资源与环境可持续发展,为政府决策分析、社会公众需求等提供互联网服务。2019年,由中国科学家倡议、13个国家组织与机构共同发起“深时数字地球”国际大科学计划,致力于搭建全球地球科学家与数据科学家合作交流的国际平台,旨在围绕地球演化这一科学命题,运用大数据、人工智能、超级计算等技术,整合过去数十亿年地球时空大数据,构建地球科学全领域知识图谱,研究生命演化、地理演化、气候演化、物质演化等重大科学问题,推动地球科学在大数据时代的创新发展^[7]。大数据时代,世界各国对地质科学的发展均采取了相应策略。其中,美国地质调查局最具代表性,制定了一系列科学战略目标:包括提高时空分辨率、展示复杂地质结构的动态变化的四维地质填图,为自然地理环境四大圈层的描述、认识提供研究技术;扩展地质调查成果应用的各项技术的应用,包括改进调查工作、数据分析、可视化和信息处理技术;基于数据整合、合成产品开发、发展交叉学科或综合学科,综合各种技术方法和手段提高信息实时性、及时性及信息覆盖面。核心目标是建立能无缝整合气候、土地、生态、能源矿产、环境、自然灾害、水资源的科学框架,将数据、技术方法、模型组织到恰当的时空尺度,深化对地球四大圈层的认识,支持国家的决策

分析^[8-10]。

1.2 国内地质信息化发展现状

2012 以来,我国政府立足信息化发展大势和国际国内大局,做出“没有信息化就没有现代化”“以信息化驱动现代化”等重大论断。面对资源与环境这两大人类生存发展的主题,充分挖掘海量数据潜在的信息,综合应用地质科学大数据,争取地质科学的新发现迫在眉睫。针对地质调查信息化建设,中国地质调查局根据《国土资源信息化“十三五”规划》,以信息化带动地质调查现代化,创建了集数据资源、应用系统于一体的“地质云”平台,搭建“汇聚-应用-回馈”的“云生态”,实现了 43 家局属单位互联互通;构建多圈层、多专业、多要素、多尺度的地球科学“一张图”大数据体系以及地质信息产品体系,实现国家层面重要原始地质资料、成果地质资料、山水林田湖草等自然资源调查数据、地质环境、土壤地球化学、地面沉降等动态监测实时数据的一体化集成与深度整合,为国家重大战略实施和社会广泛需求提供权威的地球科学数据信息服务,制定信息化技术与管理标准,强化信息化顶层设计,保障“地质云”平台稳步运行;上线“云+端”在线调查系统,基本建成地质调查“在线化”工作体系,构建地质调查信息管理系统,进一步完善“一站式”业务管理大数据系统,数据库体系更加完善,建成包括“1 个主中心+6 个区域中心+36 个专业中心”的分布式地质大数据中心,并持续更新维护核心数据库;开展基于深度学习的预填图、智能岩石矿物识别、基于遥感数据的地质灾害隐患智能识别等人工智能应用试点,持续完善深度学习模型软件平台,探索“地质大脑+人工智能”新模式^[11]。

飞速发展的新一代信息技术、深刻变革的产业格局、国家战略发展的需要、经济社会发展对地质数据信息的渴求等外部环境的发展,使得推进地质信息化工作如箭在弦,异常紧迫。本文拟就地质信息化建设及地质科学大数据应用涉及的若干问题进行初步探讨。

2 地质资料信息化服务历程

地质资料包括在地质工作中形成的文字、图表、声像等形式的原始地质资料、成果地质资料及各类标本等实物地质资料,为国民经济建设发展及进一步开展地质工作提供依据和基础。地质资料信息化服务经历了传统地质资料服务及多元化地质资料服务阶段。地

质资料包括纸质资料和电子资料两种形式。传统地质资料通过到馆查阅、办理借阅提供服务,以人工的纸质账册形式检索。随着馆藏模式的发展变化,从传统的手工信息服务过渡为多元化地质资料服务,用户可检索目录,在线浏览、下载数据,查询、阅览纸质及电子资料,同时还可享受地质资料专题加工服务、重大地质灾害地质资料应急服务、电话订阅、邮寄等服务。

2016 年以来,地质资料服务主体主要是国家地质资料馆以及省级资料馆藏机构。继“地质云”平台搭建后,地质资料服务转型升级为“地质云”平台,以地质信息服务产品的形式,提供精准化服务。

3 地质信息产品研制

地质资料是地质勘查的第一手成果,是国民经济建设、进一步开展地质工作、科学研究等的依据和基础,对科学部署找矿、减少重复工作具有重要意义。在系统分析已有地质资料的基础上,提取有用信息,可加工形成地质资料服务产品,实现地质资料信息的进一步有序化。按地质资料产品开发的层次,分为查询类产品、专题服务类产品、综合集成产品、数据挖掘类产品、数据库产品、科普类服务产品等地质信息服务产品^[12-13],实现信息产品的精准化服务。

查询类服务产品主要提供已汇交地质资料的文件级及案卷级正文、审批、附图、附表、附件、档号、编著者、形成时间、形成单位等基本信息。专题服务类产品针对某一方面的具体应用,比如找矿突破、地灾预警应急等,提取、重组、叠加成果资料数据,形成矢量、栅格或矢量栅格混合的产品。利用不同的科学方法综合分析不同阶段、不同类型的地质资料得出潜在的科学规律,累积得到的多种信息形成新的地质认识,形成综合集成类服务产品,比如综合分析重要成矿区的地层、古生物、矿床、岩石、物探、化探、遥感等地质资料,整合提取有效信息,总结规律,服务新一轮找矿突破。根据用户需求,深度挖掘多个报告中与用户需求相关的信息,形成数据挖掘类服务产品。从已有的基础、矿产、重力、自然重砂、航空物探、化探、遥感数据库中提取、聚合所需地质知识,建相关数据库,形成数据库类服务产品,缩短建库时间的同时也大大提高了数据库更新维护效率。根据国家科普导向,围绕国家重大战略部署需求,依托馆藏地学资料,以科普图书、短视频、诗歌、

化石、游戏等形式构建科普产品。

4 地质大数据存储管理与共享服务

大数据环境下实现地质数据的存储管理与共享,关键是解决海量地质大数据的分布式存储与并行计算。

地学数据的分布式存储可采用基于 Hadoop HDFS 和 Hadoop Zookeeper 的分布式文件系统 HBase。HBase 使用列式数据库,以表的形式存储数据,将非结构化数据化“散”为“整”,化“异构”为“同构”,实现地学数据稳定、高效的存储。采用包括 NameNode 节点及 DataNode 节点的 Master/Slave 架构,其中 NameNode 节点存储、维护元数据信息,DataNode 存储文件信息。NameNode 是中心服务器,一个集群中只有一个,DataNode 在集群中一般是每个节点一个,当系统容量不足时增加 DataNode 节点的数量,匹配新的服务器到整体阵列中^[14-15]。HBase 通过增加机器节点、横向扩展及自动多数据备份,有效解决地质资料的大数据量及非结构化数据的快速增长,存储数据量能达 PB 级甚至 EB 级,提高了数据容错能力与可靠性,实现了对数据的高速访问,通过数据分块、追加更新等技术手段高效存储数据,使用户能跨平台使用地质大数据计算资源。

构建并行计算架构需解决多源数据的存储与调用,以及矢量空间数据在计算节点的通信、并行计算方法。当前较流行的并行计算框架主要有 Hadoop MapReduce、Spark 等,但这些并行计算框架需要搭建大型并行计算平台,而高性能服务器价格昂贵,且要进行适合多源数据的开发。相对于上述计算框架,基于 MPI 的并行计算方法可直接使用 MPI 提供的接口编程实现,使用普通 PC 机即可完成。

并行计算架构包括数据管理层、任务调度层、通信层、计算层。

1)数据管理层:数据管理层需要在明确应用需求的基础上,根据不同的应用需求选择相应的数据模型及存储模式。传统关系型数据库 SQL 依然强大,但当 SQL 应用接近极限边缘时,可采用 NoSQL。NoSQL 包括文档数据库、图数据库、Key-Value 数据库、BigTable 类型数据库等。NoSQL 针对大数据体量支持横向扩展,有良好的可扩展性,基于动态结构,支持灵活的数

据模型,读写性能高。结构化数据主要依靠关系型数据库,将数据带进程序,采用动态分层技术虚拟存储结构化数据,把调用频率高的数据搬到最高层。非结构化数据主要采用 NoSQL 处理,将程序带进数据,使用动态归档架构把结构化数据与非结构化数据集集成到内容归档平台,通过设计软件、元数据规则、给数据加标签等建立不同维度,实现模糊查询。

2)任务调度层:任务调度层是并行计算架构的核心,主要负责将需要计算的任务分解成若干子任务,分配到各计算节点执行子任务,达到负载均衡,并整合中间结果形成最终结果。任务管理与调度采用优先队列式管理,通过 Master 节点监控集群工作状态,合理分配计算任务执行顺序。计算任务在任务提交管理器排队等待提交,等待集群资源空闲时才按任务排序依次提交,计算资源满时暂停提交。

3)通信层:使用 MPI 负责分布式计算节点间小容量数据的通信,使用 Socket 套接字进行并行计算节点间较大容量的数据传输,减少人工控制通信的开发成本。

4)计算层:计算层是整个并行计算架构的核心,结合 MPI 多进程技术从底层研发,有效提升并行计算性能。

此外,地质大数据具有时空特性,地质研究很多时候需要接入多传感器。为存储、管理高动态监测数据流,可在借鉴龚建雅所提出的实时 GIS 时空数据模型的基础上^[16],采用基于事件多因素驱动的时空数据概念模型、基于对象与事件的时空数据逻辑模型、基于系统工程库的管理结构。多因素是指地质过程中的多种相关影响因素,不同因素触发不同事件,这些事件又会影响相应对象,促使其改变当前状态。事件由事件发送队列、事件接收队列、事件池按时间顺序统一管理。时空数据存储管理逻辑模型包括地质数据时空对象、地质数据存储管理事件、地质数据存储管理时空过程。时空对象产生存储管理事件,时空对象与存储管理事件由地质数据存储管理时空过程统一管理。时空对象的运算有两方面:一是关于数据库添加、删除、修改的数据层的运算;二是地理、地质实体对象的运算。时空对象的运算将产生新的时空对象,实体对象几何形态的变化可视为产生的新的时空对象。采用系统工程库负责程序及各个过程协调运行,管理空间对象模

型、过程集、传感器监测等观测数据、事件、版本及多因素驱动模型,将地质时空对象结构分为时空对象列表和时空对象操作两部分,实现对地质时空对象的管理。地质时空对象由地质空间类和地质时间类构成。地质空间类包括地质要素和地质空间参考,地质时间类包括地质时间实体和地质时间参考。为实现模拟传感器数据接入效果,需构建 OPC 数据模拟模块。载入模拟数据后,将数据发送到 OPC 服务器端观测项。当客户端连接 OPC 模拟服务器时,通过将传感器观测项与 OPC 服务器端最基本的数据结构单元 Item 项绑定,一个传感器对象绑定一个或多个观测项,进而实现基于 OPC 接口的传感器关联,进行数据访问^[17]。

为高效管理、调度和应用地质时空大数据,还需要采用分布式并行时空索引技术。分布式并行时空索引本质上是一种本地索引,为并行索引的分布式时空索引,技术包括全局的分布式时空索引及节点内部的本地并行时空索引两部分。分布式时空索引主要的时间消耗都在网络消息的传输上,因此分布式时空索引与网络结构关系密切。分布式时空索引的根本是要缩短网络距离,减小网络消息数量,尽可能将时间邻近、空间邻近的时空对象分布在网络邻近的节点上,降低网络访问频率。本地并行时空索引与网络结构无关,尽可能让时间邻近、空间邻近的时空对象分布在同一数据块或邻近数据块,以降低外存访问效率。分布式有主从模式及对等模式两种。主从模式的查询从主节点开始,继而本地查询进一步的节点服务器,并将结果反馈给主节点,主节点反馈最终结果给查询用户。对等模式中,各个节点地位相等,查询可以从任意节点开始,继而在涉及的其他节点的本地索引中查询,查询结果根据网络负载及通信协议由指定节点反馈给查询用户。本地并行时空索引采用基于间隔关系算子的并行时空索引。基于间隔关系算子的本地并行时空索引核心思想是将时空查询转换成可并行计算的间隔关系算子,采用基于三角区域递归分解的虚拟二叉树实现各种间隔关系算子,通过并行调用虚拟二叉树实现并行时空索引^[18-19]。

5 地学知识图谱

知识图谱是数据挖掘、NLP 等 AI 技术的融合。维基百科对知识图谱的定义为借助统计学、图论、计算机

技术等手段,以可视化方式展示科学学科体系的内在结构、学科特点、研究前沿等信息的一种计量学方法。本质上,知识图谱是客观世界中存在的各种实体、概念及其关系构成的语义网络图,既呈现科学知识的变化过程,又表明知识间的结构关系,从而形式化地描述客观世界中的各类事物及其关联关系。知识图谱的出现改变了传统知识获取模式,地球科学作为典型数据密集型科学,基于知识驱动的地球科学大数据分析,地学知识图谱构建及地学知识演化能更好地整合全球地球演化数据,推动地球科学研究范式的变革^[20-21]。

地学知识图谱是以结构化的图模式全面、清晰、明确地组织相关知识及其相关关系,采用国际标准化的编码形式化表述,形成符合计算机处理模式的知识体系。地学知识图谱的构建,需要在一般知识图谱构建的基础上,融合地学多尺度时空特性^[22-23]以及地学图、文、数等数据要素。

大规模的互联网用户是网络空间大数据和各类应用程序的使用者,更是贡献者,构成了支撑大量成功应用的群智资源。为有效管理、协同利用开放网络环境下的群智资源,实现群体智能的最大化,群智协同计算应运而生,基于互联网的人类群体智能为很多领域的研究带来了创新性的求解方案。地质学是一门需要经验的学科,对岩石、矿物及各种地质构造现象的认识都需要地质专家常年积累相关知识,同时,对地质学模棱两可的地方的形式化表达极其困难。因此,通过地质专家个体基于互联网的大规模群体协同构建和演化地学领域相对成熟的事实性、概念性知识,形成知识图谱,每一个参与者把自己所知转化为相应的知识图谱片段,多个个体并行进行自由探索、自动融合、主动反馈 3 个活动,通过自动化算法拼接各个体片段信息,形成更为完整的知识图谱。同时,建立知识价值贡献评估方法及领域内专家的信誉评价模型,形成有效的群智协同激励机制,建立可持续共建共享模式。当不同专家发生意见分歧时,采用知识贡献度与可信用度评价的融合修正方法自动检测知识冲突。同时,研究和具有隐私知识保护能力进而保证知识产权的知识存储方法,避免发生专家宣称有知识但不共享的局面。基于群智协同的地学知识图谱构建框架如图 1 所示。

对地学文本文献资料如专著、论文、报告等采用文本文献资料挖掘方法构建知识图谱。首先是海量文本

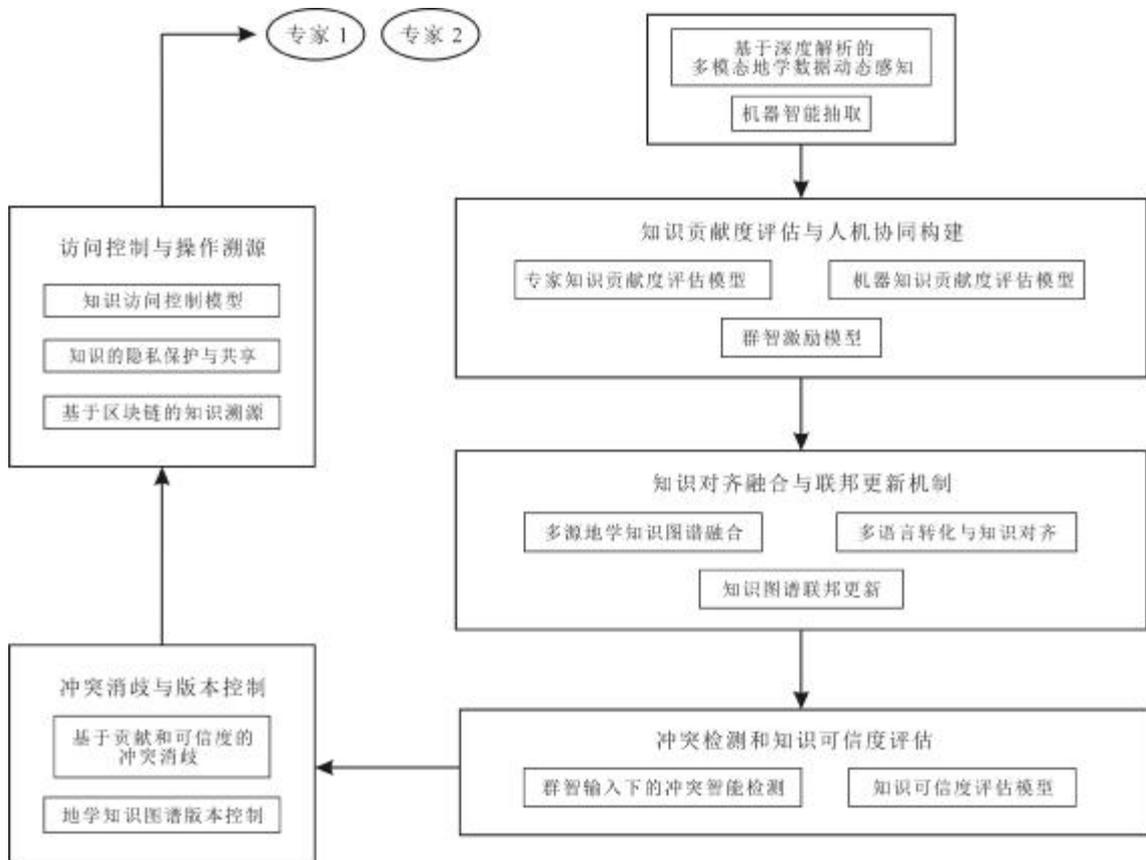


图 1 群智协同的地质知识图谱构建

Fig. 1 Construction of geoscience knowledge graph based on group intelligence cooperation

数据抽取, 基于网络爬虫采集不同数据源并对采集到的多源数据进行命名实体识别与实体关系识别, 同时, 完成冗余检测. 当文档相应的 URL 地址相同, 认为文档 100% 相似. 采用文本相似程度算法判断文档标题相似性, 采用文本摘要算法计算各文档唯一性签名, 并利用该签名计算正文内容的相似程度, 根据计算结果剔除冗余数据. 其次, 对同一来源的非结构化文本标记关联属性, 在标记的基础上, 对要处理的文本进行分词和停用词过滤、特征提取、非实质性的半结构化文本消除、纯文本提取. 关联不同来源但相似的非结构化文本, 深度分析地图符号及地图各类空间关系, 基于规则过滤及神经网络模型抽取文本关键词信息. 基于关键词出现频率与位置, 获取主题词位置, 提取主题词, 抽取相应的对象实体, 采用倒排索引实现词语-文档矩阵的具体存储, 其中与关键词相应的记录表项为该关键词的全部文档, 进而迅速找到与关键词相应的全部文档. 然后, 基于关键词图模型等深度学习完成实体对象与知识抽取. 最后, 基于实体特定的语

义时空关联特征, 计算语义相似度对实体聚类, 进而实体消歧. 若多个实体表征同一个对象, 则在多个实体间构建对齐关系, 并对实体包含的信息进行融合、聚集. 基于表示学习训练信源属性, 实现属性对齐. 基于海量地质文献资料的地质知识图谱构建如图 2 所示.

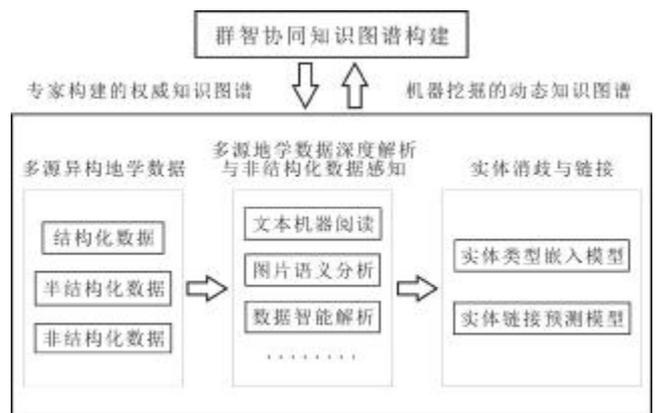


图 2 基于海量地质文献资料的知识图谱构建

Fig. 2 Construction of knowledge graph based on massive geoscience literature

当知识图谱发展到一定程度,可将群智协同以及基于海量地学文献资料的知识图谱构建相结合,实现地学知识图谱的混合型构建.地学知识图谱的广泛应用可推动地学知识推理、高精度的地质时间轴构建以及智能地图编辑与制图的发展.同时,地学知识与地球系统模型的结合可推动矿产资源预测、探测的研究^[24].

6 地质大数据可视化

地质大数据可视化是地质体空间形态结构及属性模型的建立,涉及的关键技术包括:1)地质体动态精细建模;2)海量空间数据的地质体模型与属性数据的并行可视化;3)可视化分析技术;4)虚拟现实与增强现实技术^[25].

地质体的动态精细建模涉及两个关键问题:一是智能建模,另一个是局部动态的快速更新.智能建模需要基于知识驱动、拓扑推理,在形成对研究区域地质结构、构造、同一地质时代地层、岩性、岩相等认识的基础上,结合专家知识、行业标准、语义分析技术构建专家知识库,然后基于层面、体元模型动态精细建模.层面模型并行建模需要根据拓扑推理及相关地质知识划分层面空间插值,将插值任务传给各个CPU,各个CPU完成插值任务传给主CPU,由主CPU完成合并层面模型.体元模型并行建模,面向多核CPU与计算集群新型硬件架构,采用MPI并行技术,基于八叉树结构,采用三重轴向扫描剖分算法进行体元剖分.根据栅格映射准则将所有栅格映射到八叉树结构中的相应位置,基于栅格合并准则“由下而上”递归合并所有子节点直至根节点,形成八叉树结构的多尺度栅格模型,完成体元并行构建.

最后,基于多点地质统计学随机模拟方法完成属性建模,并将属性建模模拟结果映射到体元模型网格,完成地质体模型的构建.若局部有新的钻孔等需要更新地质体模型时,采用不规则六面体网格与三角网模型进行整体地质体与局部同尺度地质体模型构建,然后采用Octree树与BSP树搜索更新范围,并利用新的需要更新的局部体元将其替换,再从新体元模型提取层面模型及构造面模型,完成层面模型的拓扑与地质属性的更新^[26-28].

对地质体属性数据场采用基于CUDA、GPU的并

行可视化技术.显卡厂商NVIDIA推出的通用并行计算架构CUDA使GPU(图形处理器)能解决复杂的计算问题,包含CUDA指令集架构(ISA)、GPU内部的并行计算引擎.GPU采用流式并行计算模型,可对每个数据进行独立的并行计算.异构计算在软件层面上都分为主机代码和设备代码,经应用分析、内存资源分配、线程资源分配到具体函数而实现.CPU常作为主机,完成串行计算.GPU作为并行处理器,完成并行计算.对于地质体大数据,可配合大数据索引机制,划分可视化任务,然后由设备完成地质体的渲染绘制,由主机对已渲染图像完成并行图像合成与结果输出.

此外,在传统三维可视化基础上,结合Unity 3D虚拟现实开发引擎以及增强现实插件Vuforia的地质虚拟现实、增强现实可视化系统,可提高多元、异构、时空、多尺度地质数据的表达能力.虚拟现实技术具有提供动态处理数据的能力,使人可以触摸数据、得到沉浸式交互的触觉体验,更容易理解、操作数据,便于从不同角度查看数据,发现新的知识.目前,这项技术已在海地震、水文地质模型、石油开发、多尺度地质数据3D沉浸式可视化等领域得以应用^[29-30].

7 区块链技术应用

区块链能提供卓越的数据安全性、数据质量、隐私性、透明性、自动化.将区块链技术应用于大数据中,使平台上的数据无法被随意更改,使大数据极具弹性,存储量增大,为大数据平台提供有力的技术补充.

区块链技术是一项革命性的技术,它最初只是作为一种数字货币交易技术,但它的潜力远不止如此.它是一种去中心化的、安全的、透明的数据库技术,可以让信息更加安全、透明、公正.区块链是无人管理的分布式数据库,所有的数据被分散存储在能相互验证并协作来维护系统安全的不同节点上.区块链去中心化的方式使得区块链技术不容易被黑客攻击,所有参与者都有更大的控制权.区块链技术最大优势是具有透明性,每一个交易都被记录在不可篡改的账本中,所有人均可查看交易记录,大大减少舞弊可能性^[31-32].区块链采用带时间戳的链式区块结构,利用哈希处理等加密算法的数据脱敏技术对交易信息加密,并将信息编码为数字与字母组成的字符串,保证数据的隐私性;采用经济激励机制、一整套协议、所有节点共同参

与的方式对添加到区块链中的新区块验证,集体维护数据;将具有动态性的、可编程的智能合约代码写在区块链上,规则公开透明、获多方承认,由事件驱动自动触发执行合同,允许在没有第三方的情况下可信交易,交易公开可见、可追踪、且不可逆转,所有交易由机器执行,避免人仲裁及作弊。区块链被保存在服务器中,这些服务器在区块链系统中即为节点。区块链上每个节点都有所有区块的完整副本,这些节点相互连接构成区块链的分布式网络,基于数字签名、非对称加密技术抵御黑客及隐私泄露。区块链包括底层技术及通证,通证使区块链与分布式数据库有本质区别,通证三要素包括数字权益证明、加密、可流通^[33-34]。

在地质领域,区块链可用于地学数据存储、地学数据标识、地学数据索引与区块链生成、地学数据确权、地学数据产权交易、地学数据维权、良好地质生态构建等领域。基于区块链将要共享的地学数据通过哈希运算生成相应的哈希树,通过开放式索引识别机制对各类地学数据建立地学数据标识,选用随机分配的共识机制将待共享的地学数据生成区块并与前期区块链成链条。利用时间戳及哈希根值为地学数据的版权所有生成独一无二的数字化身份信息进行产权确权,依托智能合约完成地学数据的产权交易。一旦存在侵权等行为,区块链系统会自动追溯到发布源头并通过智能合约进行处罚。基于通证设计地质数据所有者、共享者等成员的共赢机制,激励更多的数据所有者贡献自己的数据,志愿共享。基于区块链将新的项目策划方案及智能合约上链,感兴趣者可直接参与项目、资助项目,也可通过用户投票自动删除蹩脚项目,优质内容与低俗内容的发现用户均可获通证奖励,地质社

区的任何人都可以用地质通证的方式激励,地质通证的持有者形成全球利益共同体^[35],人人赋能,进而构建通证经济模式下良好地质商业生态,促进全球地质数据共享。基于区块链架构实现数据管理流程如图3所示。

8 结论

(1)地质信息化是推动地质工作现代化的必要手段。60多年来,地质工作与信息技术的结合推动了地质工作向前发展,并形成了地质信息化工作领域。今后,在信息技术的推动下,必将重塑地质工作理念与方法,提高地质工作质量、效率。

(2)地质科学大数据的应用涉及到多源异构数据的采集、存储、管理,数据挖掘,数据可视化以及为大数据平台提供强大支持的区块链技术等。数据挖掘有聚类分析、主成分分析、关联规则、推荐系统算法、大图形的社区结构识别、机器学习与深度学习等多种方法。限于篇幅,文中只讨论了基于群智协同以及文本文献资料地学知识图谱的构建。知识图谱是人工智能的重要基础设施,突破了传统的数据存储、数据使用方法,以图结构呈现知识,是近年来人工智能与交叉学科领域的研究热点。地学知识图谱研究尚处于起步阶段,目前,地学动态知识图谱的构建多是基于文献资料,知识表达多采用“节点-边”的基本架构。如何从更多来源的数据中发现知识,如何检测新知识的可靠性,如何快速更新知识图谱,如何融合各类地质图件生成知识图谱等尚未实现。

(3)地质大数据应用处理的数据是全体数据而不是抽样数据,揭示的数据内涵是相关关系,而不是因果

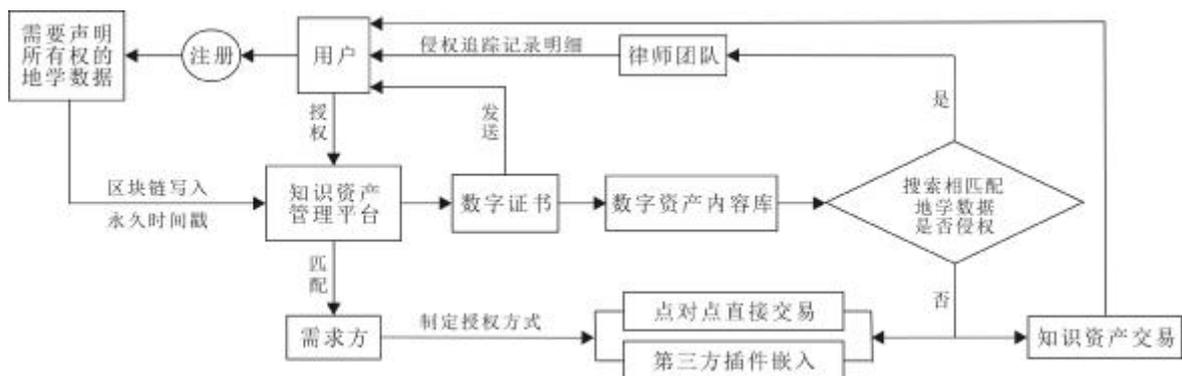


图3 区块链视角下地学数据管理流程图

Fig. 3 Flowchart of geoscience data management from the perspective of blockchain

关系,更多的是追求数据而不是算法,大数据的简单计算比抽样的复杂计算更有效.基于大数据的密集型计算方式,即科学研究的第四范式,突破了采样随机性和样本空间狭小等限制,能基于数据库实现,对地质资源、地灾、环境的预测有重大意义.

参考文献(References):

- [1] Esterle J S, Carr G R. The glass earth [R]. Australian Institute of Geoscientists News, 2003: 1-6.
- [2] Stafleu J, Maljers D, Busschers F S, et al. TNO — Geological survey of the Netherlands: 3-D geological modeling of the upper 500 to 1000 meters of the Dutch subsurface [M]//Berg R C, Mathers S J, Kessler H, et al. Synopsis of current three-dimensional geological mapping and modeling in geological survey organizations. Illinois State Geological Survey, 2011: 64-68.
- [3] Russell H A J, Boisvert E, Logan C, et al. Geological survey of Canada: Three-dimensional geological mapping for groundwater applications [M]//Berg R C, Mathers S J, Kessler H, et al. Synopsis of current three dimensional geological mapping and modeling in geological survey organizations. Illinois State Geological Survey Circular 578, 2011: 31-41.
- [4] Mathers S. 3D geological mapping (modeling) in geological survey organisations and the new British geological survey initiative to build a national geological model of the UK [C]//Three-dimensional workshops for 2011. Minneapolis, Minnesota: Geological Survey of Canada.
- [5] Castagnac C, Tmffert C, Bourguin B, et al. French Geological Survey (Bureau de Recherches Géologiques et Minières): Multiple software packages for addressing geological complexities [M]//Berg R C, Mathers S J, Kessler H, et al. Synopsis of current three dimensional geological mapping and modeling in geological survey organizations. Illinois State Geological Survey Circular 578, 2011: 32-47.
- [6] Jacobsen L J, Glynn P D, Phelps G A, et al. U.S. Geological Survey: A synopsis of three-dimensional modeling [M]//Berg R C, Mathers S J, Kessler H, et al. Synopsis of current three-dimensional geological mapping and modeling in geological survey organizations. Illinois State Geological Survey, 2011: 69-79.
- [7] 王学求. “化学地球”国际大科学计划取得重要进展 [J]. 中国地质, 2018, 45(5): 858.
Wang X Q. Significant progress achieved in the International Major Science Program of “Chemical Earth” [J]. Geology in China, 2018, 45(5): 858. (in Chinese)
- [8] Water Resources Division of U.S. Geological Survey. Facing tomorrow’s challenges: U.S. Geological Survey science in the decade 2007-2017 [R]. Reston, Virginia: U.S. Geological Survey, 2007.
- [9] Linda C S, Belnap J, Goldhaber M, et al. Geology for a changing world 2010-2020: implementing the U.S. Geological Survey science strategy [R]. Reston, Virginia: U.S. Geological Survey, 2011.
- [10] Bristol R S, Euliss N H, Booth N, et al. Science strategy for core science systems in the U.S. Geological Survey, 2013-2023 [R]. Reston, Virginia: U.S. Geological Survey, 2012.
- [11] 中国地质调查局. 中国地质调查局年鉴 [M]. 北京: 地质出版社, 2022: 115-117.
China Geological Survey. Yearbook of China Geological Survey [M]. Beijing: Geological Publishing House, 2022: 115-117.
- [12] 高学正, 李晓蕾, 等. 地质资料网络服务产品及服务价值研究 [J]. 中国矿业, 2017, 26(4): 64-68.
Gao X Z, Li X L. Research on the product and service value of geological data network [J]. China Mining Magazine, 2017, 26(4): 64-68.
- [13] 于清海, 丁克永, 张明超, 等. 基于地质资料的地质科普发展探讨 [J]. 中国矿业, 2018, 27(8): 168-171.
Yu Q H, Ding K Y, Zhang M C, et al. Discussion on the development of geoscience popularization based on geological data [J]. China Mining Magazine, 2018, 27(8): 168-171.
- [14] 李婧, 陈建平, 王翔. 地质大数据存储技术 [J]. 地质通报, 2015, 34(8): 1589-1594.
Li J, Chen J P, Wang X. A study of the storage technology of geological big data [J]. Geological Bulletin of China, 2015, 34(8): 1589-1594.
- [15] 魏东琦, 江宝得, 张静雅. 非结构化地质数据内容存储方法研究 [J]. 西北地质, 2021, 54(4): 266-273.
Wei D Q, Jiang B D, Zhang J Y. Research on content storage method of unstructured geological data [J]. Northwestern Geology, 2021, 54(4): 266-273.
- [16] 龚建雅, 李小龙, 吴华意. 实时 GIS 时空数据模型 [J]. 测绘学报, 2014, 43(3): 226-232, 275.
Gong J Y, Li X L, Wu H Y. Spatiotemporal data model for real-time GIS [J]. Acta Geodaetica et Cartographica Sinica, 2014, 43(3): 226-232, 275.
- [17] 刘刚, 吴冲龙, 何珍文, 等. 面向地质时空大数据表达与存储管理的数据模型研究 [J]. 地质科技通报, 2020, 39(1): 164-174.
Liu G, Wu C L, He Z W, et al. Data model for geological spatiotemporal big data expression and storage management [J]. Bulletin of Geological Science and Technology, 2020, 39(1): 164-174.
- [18] 郑祖芳. 分布式并行时空索引技术研究 [D]. 武汉: 中国地质大学 (武汉), 2014.
Zheng Z F. Distributed parallel spatio-temporal index techniques [D]. Wuhan: China University of Geosciences (Wuhan), 2014.
- [19] 田善君. 面向地质大数据存储管理的时空数据模型研究 [D]. 武汉: 中国地质大学, 2016.
Tian S J. Spatiotemporal data model for geological big data storage management [D]. Wuhan: China University of Geosciences, 2016.
- [20] 张旗, 周永章. 大数据正在引发地球科学领域一场深刻的革

- 命——《地质科学》2017 年大数据专题代序[J]. 地质科学, 2017, 52(3): 637-648.
- Zhang Q, Zhou Y Z. Big data will lead to a profound revolution in the field of geological science[J]. Chinese Journal of Geology, 2017, 52(3): 637-648.
- [21]张博, 康奥. 国内地质公园研究演化趋势与热点分析——基于知识图谱视角[J]. 地质论评, 2021, 67(2): 557-569.
- Zhang B, Kang A. Research highlight and evolution trend analysis of geoparks in China: Based on the perspective of knowledge mapping [J]. Geological Review, 2021, 67(2): 557-569.
- [22]张雪英, 张春菊, 吴明光, 等. 顾及时空特征的地理知识图谱构建方法[J]. 中国科学: 信息科学, 2020, 50(7): 1019-1032.
- Zhang X Y, Zhang C J, Wu M G, et al. Spatiotemporal features based geographical knowledge graph construction[J]. Scientia Sinica Informationis, 2020, 50(7): 1019-1032.
- [23]王鑫, 邹磊, 王朝坤, 等. 知识图谱数据管理研究综述[J]. 软件学报, 2019, 30(7): 2139-2174.
- Wang X, Zou L, WANG C K, et al. Research on knowledge graph data management: A survey[J]. Journal of Software, 2019, 30(7): 2139-2174.
- [24]周成虎, 王华, 王成善, 等. 大数据时代的地学知识图谱研究[J]. 中国科学: 地球科学, 2021, 51(7): 1070-1079.
- Zhou C H, Wang H, Wang C S, et al. Geoscience knowledge graph in the big data era[J]. Science China Earth Sciences, 2021, 64(7): 1105-1114.
- [25]王元昊, 高振记, 宋越. 三维地质模型质量评估方法研究进展综述 [J]. 华北地质, 2023, 46(1): 80-86.
- Wang Y H, Gao Z J, Song Y. Review of research progress on quality assessment methods of 3D geological models[J]. North China Geology, 2023, 46(1): 80-86.
- [26]江锦成, 郭甲腾, 吴立新, 等. 三维地学实体多粒度栅格剖分与布尔运算的并行算法[J]. 科技导报, 2011, 29(35): 18-23.
- Jiang J C, Guo J T, Wu L X, et al. Multi-granularity rasterization and Boolean operation parallel algorithm for 3D spatial entities [J]. Science & Technology Review, 2011, 29(35): 18-23.
- [27]田宜平, 吴冲龙, 翁正平, 等. 地质大数据可视化关键技术探讨 [J]. 地质科技通报, 2020, 39(4): 29-36.
- Tian Y P, Wu C L, Weng Z P, et al. Key technologies of geological big data visualization[J]. Bulletin of Geological Science and Technology, 2020, 39(4): 29-36.
- [28]刘杰, 赵侃. 基于 SKUA-GOCAD 的天津市潘庄凸起区域三维热储建模研究[J]. 华北地质, 2023, 46(2): 17-23, 30.
- Liu J, Zhao K. Three-dimensional modeling research of geothermal reservoirs on Panzhuang uplift area in Tianjin based on SKUA-GOCAD platform[J]. North China Geology, 2023, 46(2): 17-23, 30.
- [29]Cowgill E, Bernardin T S, Oskin M E, et al. Interactive terrain visualization enables virtual field work during rapid scientific response to the 2010 Haiti earthquake[J]. Geosphere, 2012, 84(4): 787-804.
- [30]郭艳军, 张进江, 陈斌, 等. 基于 VR 技术的多尺度地质数据 3D 沉浸式可视化与交互方法[J]. 地学前缘, 2019, 26(4): 146-158.
- Guo Y J, Zhang J J, Chen B, et al. Method development of 3D immersive visualization and interaction of multi-scale geological data based on VR technology[J]. Earth Science Frontiers, 2019, 26(4): 146-158.
- [31]郝世博, 徐文哲, 唐正韵. 科学数据共享区块链模型及实现机理研究[J]. 情报理论与实践, 2018, 41(11): 57-62.
- Hao S B, Xu W Z, Tang Z Y, et al. Block chain model of scientific data sharing and its realization mechanism [J]. Information Studies: Theory & Application, 2018, 41(11): 57-62.
- [32]郑阳, 杜荣. 区块链技术在数字知识资产管理中的应用[J]. 出版科学, 2018, 26(3): 97-104.
- Zheng Y, Du R. Application of blockchain technology in digital knowledge assets management[J]. Publishing Journal, 2018, 26(3): 97-104.
- [33]李敏, 傅洁, 陈安蜀, 等. 面向知识服务的地质资料管理转型研究 [J]. 地质与资源, 2021, 30(1): 92-98.
- Li M, Fu J, Chen A S, et al. Study on the transformation of knowledge service-oriented geological data management[J]. Geology and Resources, 2021, 30(1): 92-98.
- [34]袁勇, 王飞跃. 区块链技术发展现状与展望[J]. 自动化学报, 2016, 42(4): 481-494.
- Yuan Y, Wang F Y. Blockchain: The state of the art and future trends[J]. Acta Automatica Sinica, 2016, 42(4): 481-494.
- [35]周永章, 刘楠, 陈川, 等. 开启区块链地质应用新时代[J]. 地质通报, 2020, 39(1): 1-6.
- Zhou Y Z, Liu N, Chen C, et al. A new geological era initiated from blockchain's application[J]. Geological Bulletin of China, 2020, 39(1): 1-6.