**中图分类号：TP39；TV882.1**采用《中国图书馆分类法》（第五版）进行分类，请补充。

**文献标识码：**A **doi：**10.3969/j.issn.1000-1379.2023.

**引用格式**：[J].人民黄河，2023,45（）：、

收稿日期：2025-02-19 修回日期：2025-03-14

基金项目：“十四五”国家重点研发计划项目（2022YFC3044400）

作者简介：吴廷鑫（1998—），男，贵州遵义人，工程师，硕士，主要从事GIS研究工作

通信作者：刘心怡（1987—），女，贵州毕节人，高级工程师，硕士，主要从事地理信息技术工程应用工作

E-mail：[liuxinyi\_gyy@powerchina.cn](mailto:liuxinyi_gyy@powerchina.cn)

大语言模型赋能数字孪生水利知识平台建设研究

杨 洋，金 玉，刘心怡

（中国电建集团贵阳勘测设计研究院有限公司，贵州 贵阳 550081）

摘 要：针对单一灾种的自然灾害知识图谱覆盖信息面窄，难以从海量复杂的信息中提炼出知识这一问题，提出一种面向灾害损失评估的黄河流域自然灾害知识图谱构建方法。自然灾害知识图谱由数据资源层、知识抽取层（包含模式层、数据层）和应用服务层构成。以自然灾害事件本体、基础地理信息本体与灾害损失评估本体模型为主体，采用自顶向下的方法建立图谱模式层。收集多源异构的自然灾害事件数据、基础地理信息数据与灾害损失数据，采用自底向上方法构建数据层，实现从多源异构数据的知识抽取、知识融合与知识存储。应用成果表明，该图谱支持时空关系的高效检索与区域并发灾害的快速识别，对1981—2018年黄河流域发生的160次重大灾害事件分析发现，灾害发生频率逐年递增，其中洪水为最常见的灾害类型，并识别出37次并发灾害，对提升黄河流域科学决策能力具有重要价值。

关键词：自然灾害知识图谱；模式层；数据层；基础地理信息；灾害损失评估；黄河流域

中图分类号：TP39；TV882.1 文献标志码：A **doi：**10.3969/j.issn.1000-1379.2025

**Construction of Yellow River Basin Natural Disaster Knowledge Graph for Disaster Impact Assessment**

WU Tingxin, Huang Rui, YANG Yang, LIU Xinyi

(Powerchina Guiyang Engineering Corporation Limited, Guiyang 550081, China)

**Abstract:** In response to the problem that natural disaster knowledge graphs for single disaster types have a narrow coverage of information and struggle to extract knowledge from vast amounts of complex data, this study proposes a method for constructing a natural disaster knowledge graph for the Yellow River Basin oriented toward loss assessment. The natural disaster knowledge graph consists of a data resource layer, a knowledge extraction layer (including a schema layer and a data layer), and an application service layer. The schema layer is constructed using a top-down approach, centered on ontological models of natural disaster events, fundamental geographic information, and disaster loss assessment. Multi-source heterogeneous data, including natural disaster event data, basic geographic information data, and disaster loss data, are collected, and the data layer is built using a bottom-up approach, enabling knowledge extraction, knowledge fusion, and knowledge storage from these diverse data sources. Application results show that the knowledge graph supports efficient spatiotemporal relationship queries and rapid identification of regional concurrent disasters. Analysis of 160 major disaster events in the Yellow River Basin from 1981 to 2018 reveals an increasing frequency of disasters over time, with floods being the most common disaster type. Additionally, 37 instances of concurrent disasters were identified. These findings demonstrate the significant value of the knowledge graph for enhancing scientific decision-making capabilities in the Yellow River Basin.

**Key words:** natural disaster knowledge graph; pattern layer; data layer; basic geographic information; disaster loss assessment; Yellow River basin

0 引言

受全球气候变化和人类活动的影响，黄河流域各类自然灾害呈现发生频次高、覆盖范围广、后果严重的特点，对人民生命和财产安全造成极大威胁[1]。高效整合黄河流域水文、地质与气象等灾害发生相关信息，为自然灾害数据分析与决策提供科学支持具有非常重要的作用。

知识图谱（Knowledge Graph）的概念于2012年由谷歌首次提出，其本质上是一种以图结构形式表示和组织知识的语义网络，网络中的节点代表实体或概念，边代表实体或概念的属性或者彼此之间的语义关系[2-4]。知识图谱按照应用领域可划分为通用知识图谱与领域知识图谱两类，其中：通用知识图谱面向通用领域，主要包含大量常识性知识；领域知识图谱又称为行业知识图谱或垂直知识图谱，是由某一特定领域专业数据构成的知识库[5-6]。知识图谱能够有效整合多源异构数据并揭示事物的内在规律，其在医疗[7]、电力[8]、农业[9]、自然灾害[10]、智能制造[11]等领域都得到了广泛应用。在自然灾害领域，知识图谱相关研究主要聚焦于单一灾种的本体建模与知识体系构建，如邱芹军等[12]提出了面向灾害应急响应的地质灾害链知识图谱构建方法，沈伟豪等[13]提出了基于多模态数据的洪涝灾害知识图谱构建方法，王益鹏等[14]提出了顾及时空过程的台风灾害事件知识图谱表示方法。黄河流域自然灾害信息来源广泛、结构不统一、时空演化机理复杂，单一灾种的自然灾害知识图谱覆盖信息面窄，难以从海量复杂的信息中提炼出知识。

针对上述问题，笔者以黄河流域自然灾害为核心，围绕自然灾害事件本体模型、基础地理信息本体模型与灾害损失评估本体模型为主体，建立图谱概念模型以及要素之间语义关联关系，实现从多源异构数据中的知识抽取、融合与存储，构建面向灾害损失评估的黄河流域自然灾害知识图谱，以期为灾害损失评估提供科学依据和技术手段。

1 研究区域概况

黄河流域西起巴颜喀拉山，流经青海、四川、甘肃、宁夏、内蒙古、陕西、山西、河南、山东九省（区），东临渤海，从西到东横跨青藏高原、内蒙古高原、黄土高原和黄淮海平原4个地貌单元，东西长度约1 900 km，流域北抵阴山，南靠秦岭，南北跨度约1 100 km，流域经度范围为95°53'—119°05'E，纬度范围为32°10'—41°50'N，流域面积达79.47万km2。黄河流域在我国经济社会发展和生态安全方面具有十分重要的地位，但自然灾害特别是水害频发，给沿岸百姓带来了深重灾难。黄河流域地理跨度大、地质条件复杂，重大自然灾害主要指上游巨型滑坡灾害、中游水土灾害和下游巨型洪灾[15]，灾害分布如图1所示。黄河流域上游受青藏高原持续隆升影响，地质构造活跃，沿河分布了大量巨型滑坡、堰塞湖、泥石流等地质灾害，历史上部分滑坡曾堵塞黄河，溃决形成超大洪水[16]。流域中游流经黄土高原，生态环境脆弱，导致水土灾害分布广、类型多、突发性强，且灾害往往以链的形式发生，严重威胁人居安全与社会稳定[17]。流域下游是华北平原，其洪水灾害频发主要受泥沙淤积、季风气候暴雨集中以及人类活动等多重因素共同作用。



图1 黄河流域重大自然灾害分布

Fig.1 Distribution of Major Natural Disasters in the Yellow River Basin

2 自然灾害知识图谱整体框架

面向灾害损失评估的黄河流域自然灾害知识图谱整体框架见图2，主要分为数据资源层、知识抽取层和应用服务层。数据资源层包括灾害事件数据、基础地理信息数据、灾害损失数据这3类原始数据，其中：灾害事件数据主要包括全球地理编码灾害数据集（GDIS）[18]、紧急灾难数据库（EM-DAT）[19]、青藏高原数据中心提供的黄河流域洪水与地震灾害数据等；基础地理信息数据主要包括行政区划、水系等数据，采用全国地理信息资源目录服务系统发布的部分公众版基础地理信息数据；灾害损失数据主要包括各年份的水资源公报、各省份统计年鉴数据，涵盖经济损失、人员伤亡、基础设施破坏等详细信息，以及黄河水文网提供的黄河水资源公报、泥沙公报与水土保持公报等。知识抽取层主要对实体概念模式进行定义和数据抽取，从而为应用服务层提供专业的知识支撑；应用服务层对外提供基于时空关系的知识检索、并发灾害识别与影响评估等应用。

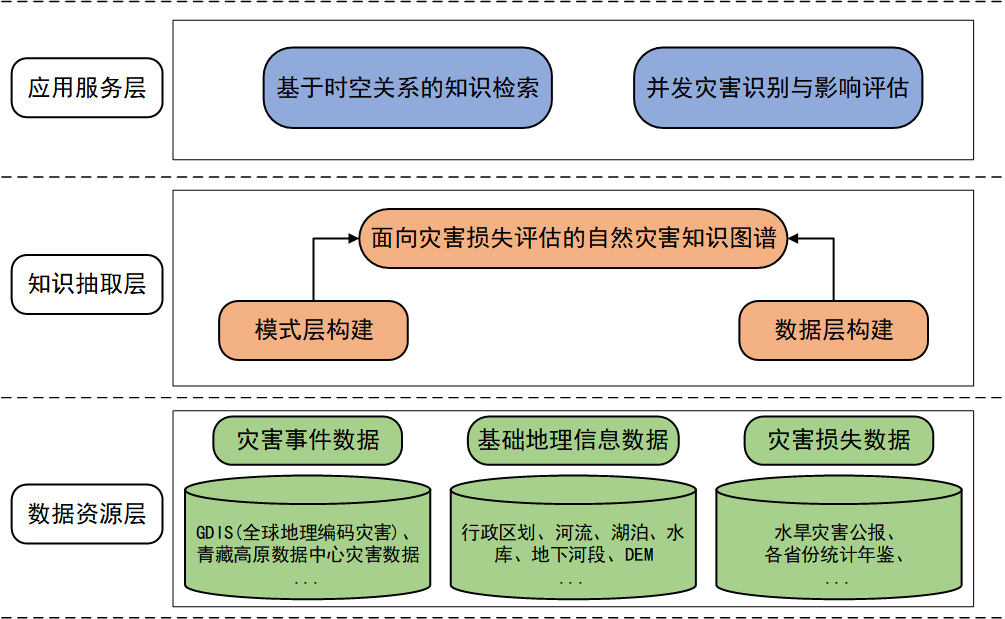


图2 自然灾害知识图谱整体框架

Fig.2 Overall Framework of Natural Disaster Knowledge Graph

3 图谱知识抽取层构建流程与方法

3.1 构建流程

自然灾害知识图谱知识抽取层构建分为模式层构建与数据层构建，具体流程如图3所示。模式层以流域自然灾害损失评估需求为出发点，自顶向下定义图谱相关实体及其属性、语义关系和约束规则等，构建面向灾害损失评估的本体库。数据层基于海量多源异构的灾害数据，自底向上识别抽取本体库中定义的实体、属性及其之间的关系，通过知识融合进行数据消歧，利用PostgreSQL与Neo4j进行存储和管理。

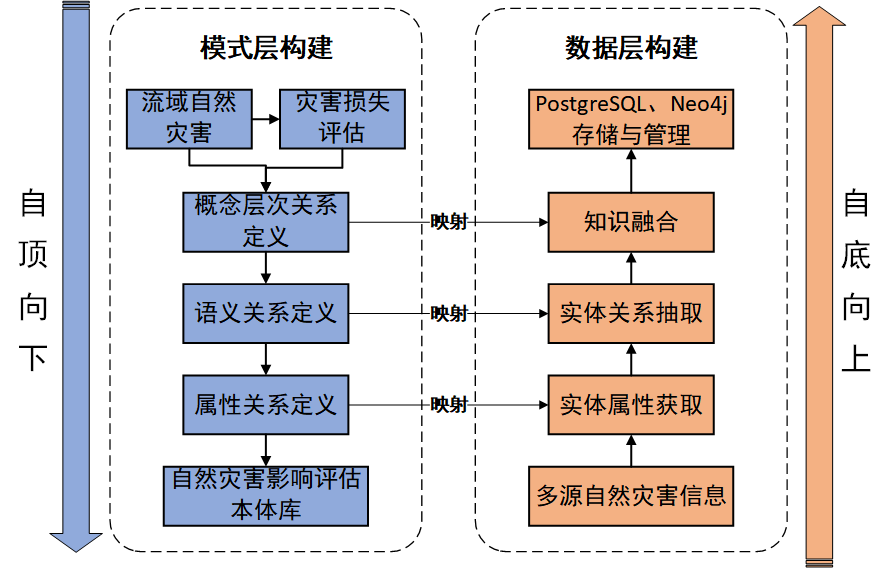


图3 知识抽取层构建流程

Fig.3 Construction Process of Knowledge Extraction Layer

3.2 模式层构建

各灾种灾害损失数据作为社会经济统计数据的重要组成部分，通常是以各级行政区为单位的汇总数据[20]，基于该特点，本研究的模式层分为自然灾害事件本体模型、基础地理信息本体模型与灾害损失评估本体模型，其中基础地理信息本体模型中的行政区划实体是另外两者的关键纽带。模式层组成如图4所示。

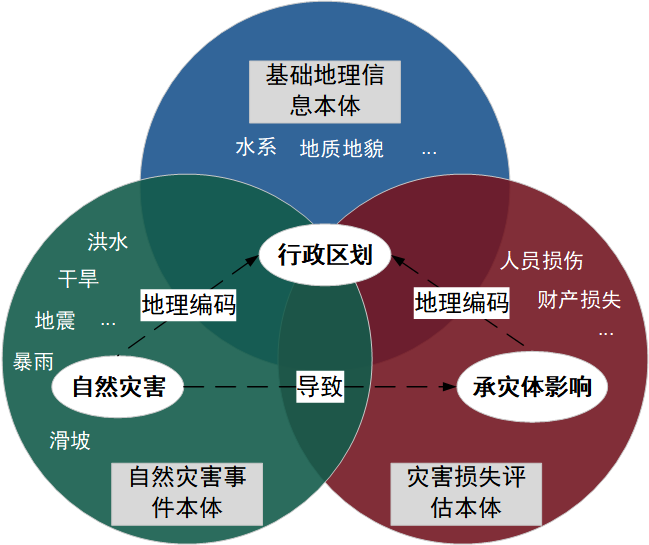


图4 模式层组成

Fig.4 Composition of Pattern Layer

自然灾害事件本体模型基于《自然灾害分类与代码》（GB/T 28921—2012）标准、以及GDIS与EM-DAT的分类体系构建而成，其主要分为水文、地质、气象三类主灾害，每一类又划分为不同子灾害，子灾害必须具备时间、地点（地名）等属性，灾害类型具体见表1。对灾害事件划分为重大灾害事件与一般灾害事件，重大灾害事件须满足以下3个条件中的任意1个[21]：有10人或以上人口因灾死亡；有100人或以上人口受到灾害影响；当地政府针对灾害事件宣布了紧急状态或请求国际援助。

**表1 灾害类型**

**Tab.1 Disaster Type**

|  |  |
| --- | --- |
| 主灾害类型 | 子灾害类型 |
| 水文灾害 | 洪水、暴雨、风暴潮 |
| 地质灾害 | 地震、滑坡、干性物质运动 |
| 气象灾害 | 干旱、极端温度、风暴 |

基础地理信息本体模型以全国地理信息资源目录服务系统发布的部分公众版基础地理信息数据为主，涵盖行政区划、水系（面、线、点）、水系附属设施（面、线、点）以及地貌与土质（面）等关键要素，构建了一个多层次的地理空间框架，部分基础地理信息数据见表2。其中行政区划实体依据《中华人民共和国行政区划代码》（GB/T 2260—2007），采用分级编码体系，实现了省、市、县、乡（镇）四级行政单元的空间拓扑关系表达与属性关联，为灾害数据的集成、共享与智能分析提供了标准化的语义框架支撑。

**表2 部分基础地理信息数据**

**Tab.2 Some Basic Geographic Information Data**

|  |  |  |
| --- | --- | --- |
| 要素名称 | 要素分类 | 主要要素内容 |
| 境界与政区 | 行政区划 | 各级行政区 |
| 水系 | 水系（面） | 湖泊、水库、双线的河流和沟渠等 |
| 水系（线） | 单线的河流、沟渠，河流结构线等 |
| 水系（点） | 泉、井等 |
| 水系附属设施（面） | 干出滩、滩涂、礁石等 |
| 水系附属设施（线） | 干出线、潮水沟、高水界、堤、闸、坝等 |
| 水系附属设施（点） | 地下河段出入口、涵洞、礁石、闸、坝等 |
| 地貌土质 | 地貌与土质（面） | 沙地、冰雪地等 |

灾害损失评估本体模型通常分为直接灾害损失与间接灾害损失2个部分。直接灾害损失通过行政区划关联的多维度数据（如人员伤亡、财产损失、房屋损毁等）可以直接量化，从而得到较为明确的灾害损失结果。但这些直接损失数据仅反映了部分灾害影响，各种自然灾害在造成直接损失的同时，也会造成巨大的间接灾害损失[21-22]。间接灾害损失通常具有滞后性、复杂性和广泛性，难以通过传统的直接灾害损失评估方法全面捕捉和量化，本研究通过为实体添加时间区间（hasTime）与空间多边形（hasLocation）属性，实现灾害影响的时空约束推理。例如，当某区域发生持续干旱（DroughtEvent，时间范围为2023年6—9月），系统自动检索该时段内空间重叠的农业实体（Farm），并沿农业灾害事件（cropFailure）→粮食价格变动（foodPriceRise）→社会稳定性（socialUnrest）路径推导社会稳定性风险。

在构建面向灾害损失评估的黄河流域自然灾害知识图谱过程中，各类实体之间基于其定义和内在机理，存在复杂的语义关系，主要包括以下3类：1)层次关系，用于表达实体之间的从属或包含关系，如灾害类型、行政区划的层级结构等；2)关联关系，用于表达不同实体之间的相互联系，如灾害事件与行政区划相互关联、不同灾害事件之间相互关联；3)空间关系，大部分实体都具备空间信息，本文主要考虑了GIS空间关系九交模型的方向关系与拓扑关系[23]。

3.3 数据层构建

数据层构建涵盖从结构化、半结构化到非结构化数据的知识抽取、知识融合与知识存储的全过程，具体如图5所示。

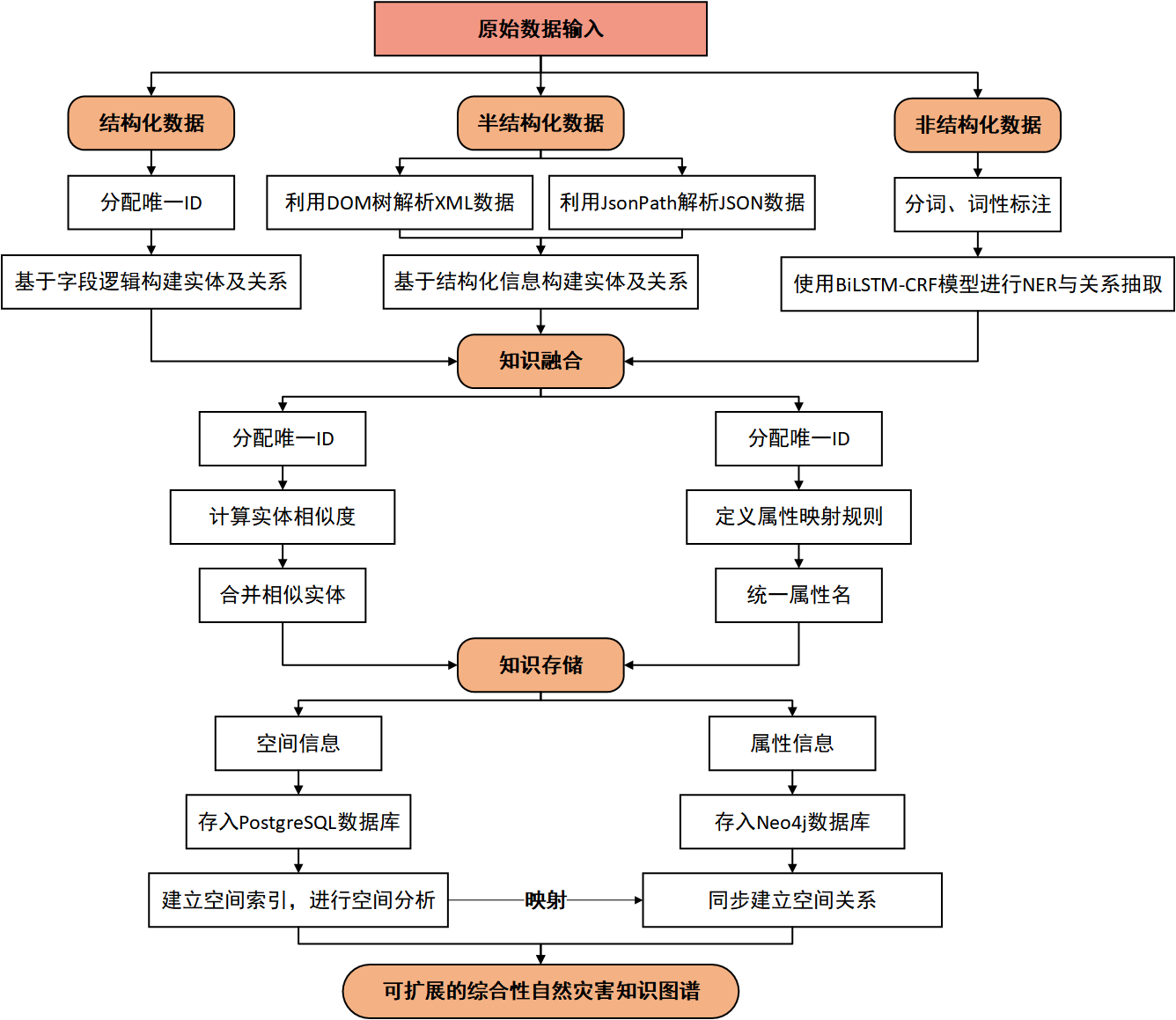


图5 数据层构建流程

Fig.5 Data Layer Construction Process

在知识抽取阶段，各类数据被转换为“实体—关系—实体”的三元组形式。对于结构化数据（主要是二维表结构数据），每一行代表一个独立的实体，通过为其分配唯一标识符，并基于字段间的逻辑关系构建相应的实体及关联管理。半结构化数据（主要是XML、JSON等格式的树状层次结构数据）通过基于DOM树的XML解析器和JsonPath进行结构化解析，然后基于解析后的结构化数据构建实体及关系。非结构化数据（主要是统计年鉴等PDF文件）的知识获取是本图谱构建过程中的关键技术与难点，具体如图6所示，首先，利用光学字符识别（OCR）方法识别各类统计年鉴、水资源公报的文本内容；其次，通过jieba库进行分词与词性标注；再次，采用双向长短期记忆网络(BiLSTM)与条件随机场(CRF)的联合模型（BiLSTM-CRF）进行命名实体识别与关系抽取，BiLSTM能够有效捕获长距离上下文依赖关系，CRF层对标签序列间的约束关系进行建模，确保输出的标签序列满足语法结构约束；最后，基于实体类型与句法分析将抽取的实体和关系转化为“头实体-关系-尾实体”的三元组结构。

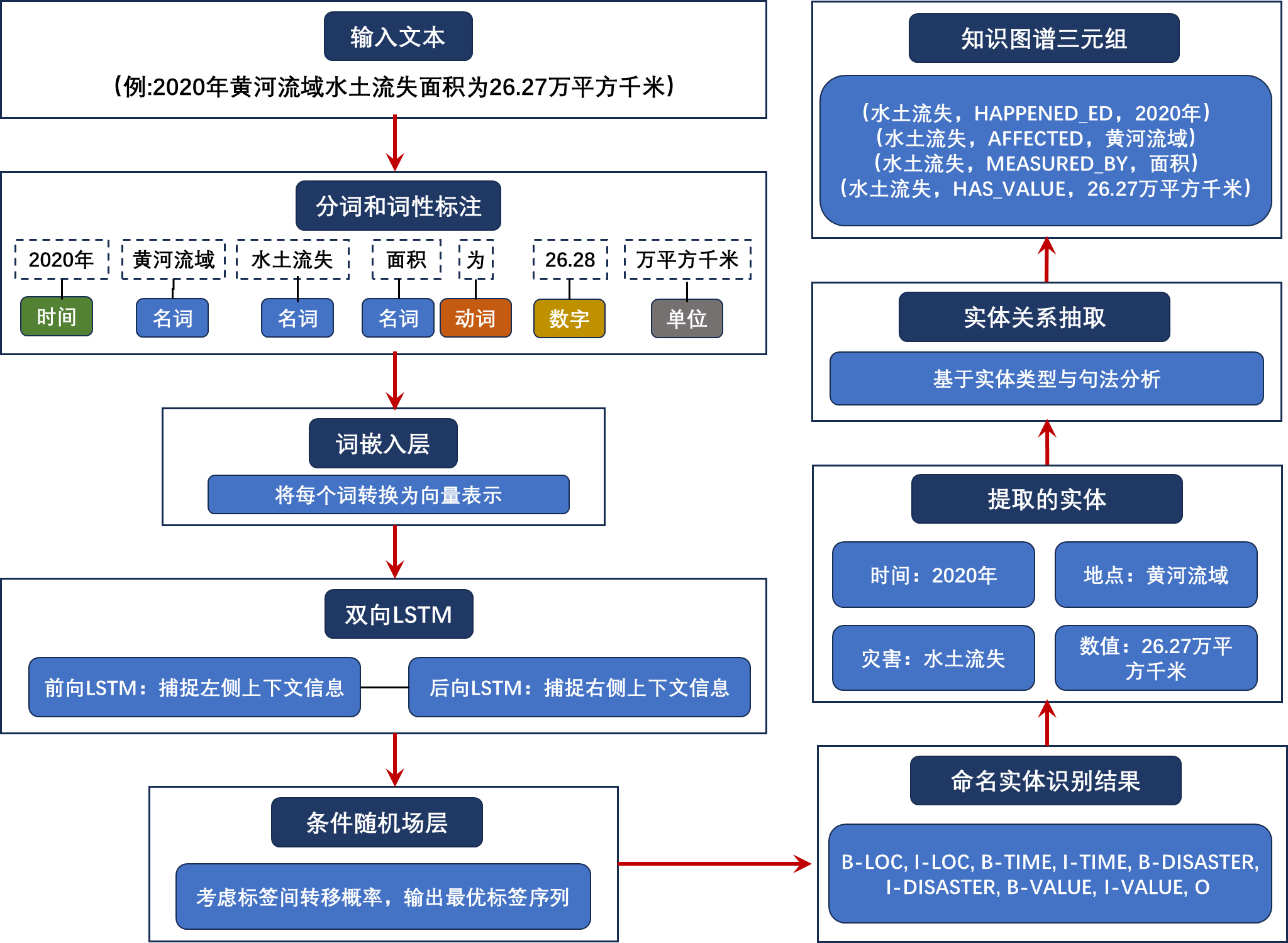


图6 BiLSTM-CRF模型命名实体识别原理

Fig.6 Principle of Named Entity Recognition in BiLSTM-CRF Model

知识融合旨在消除异构数据源之间的差异，实现信息的一致性整合。本研究采用实体对齐和属性对齐两种策略来达成这一目标。实体对齐过程中，对实体间的相似度进行计算，当相似度超过预设阈值时，将多个实体合并为单一实体。属性对齐依赖于定义明确的属性映射规则，确保来自不同数据源的属性能够统一到一致的属性名下。

完成上述步骤后，将所有实体的空间信息存储于PostgreSQL数据库并建立空间索引，属性信息存入Neo4j数据库形成实体节点。每个对象在2个数据库中共享唯一ID，建立PostgreSQL空间表与Neo4j节点间的双向映射机制。基于PostgreSQL数据库的PostGIS插件提供丰富的空间函数和操作符（如ST\_Intersects、ST\_Buffer），实现不同对象之间的空间关系提取，同时利用Neo4j数据库的Cypher查询语言实现多维语义关系的动态推理与知识发现，最终构建一个既具备强大空间分析能力又拥有丰富语义关联的综合性自然灾害知识图谱。

4 自然灾害知识图谱具体应用

传统的知识图谱通常侧重于表达语义关系，而忽略了对象之间的时空属性特征。本研究将实体空间信息存储在支持空间索引与空间分析运算的PostgreSQL数据库中，结合相关算法与函数，半自动化地构建了实体之间的时空关系并导入Neo4j，从而支持时空关系的知识检索。通过对灾害事件实体时间与空间属性的设置，可快速识别区域并发灾害，同时通过行政区划关联的灾害影响数据，支持高效获取灾害损失评估结果。

4.1 基于时空关系的知识检索

以黄河上游（河源至内蒙古河口镇）、中游（河口镇至河南省桃花峪）和下游（桃花峪至入海口）为研究区域[24]，将黄河流域上、中、下游分区分别作为单独的实体导入知识图谱，通过PostgreSQL的ST\_Contains空间分析函数，建立黄河流域上、中、下游分区与1981—2018年重大灾害事件（包括洪水、风暴和地震）的时空包含关系，实现灾害数据的结构化语义建模，并导入Neo4j进行存储，黄河流域上中下游分区重大灾害事件数据具体如图7所示。该知识图谱采用PostgreSQL与Neo4j的双重存储架构，既保留空间拓扑关系又支持语义推理，可有效突破传统GIS系统在跨时空维度关联分析中的局限。根据图谱关系统计可得，黄河流域从1981—2018年共发生了160次重大灾害事件，灾害类型主要是洪水（80次）、风暴（31次）和地震（19次），黄河上游区域共发生灾害64次，中游区域共发生灾害96次，下游区域未发生重大灾害，洪水都是各分区最为频繁的灾害类型，且每个区域的灾害事件发生次数都呈逐年增加趋势，具体如图8所示。

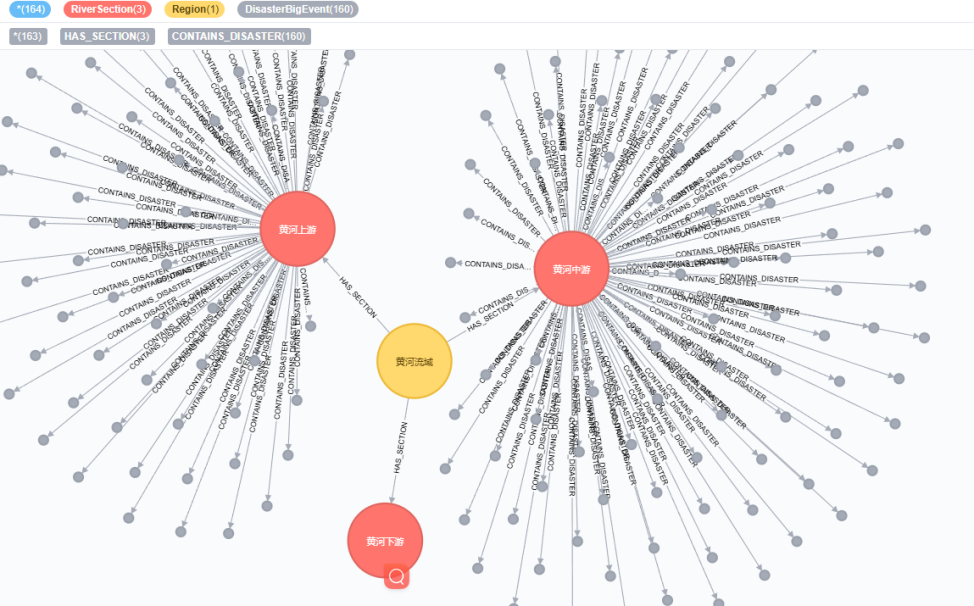


图7 黄河流域上中下游分区重大灾害事件数据

Fig.7 Data of Major Disaster Events in the Upper, Middle and Lower Reaches of Yellow River

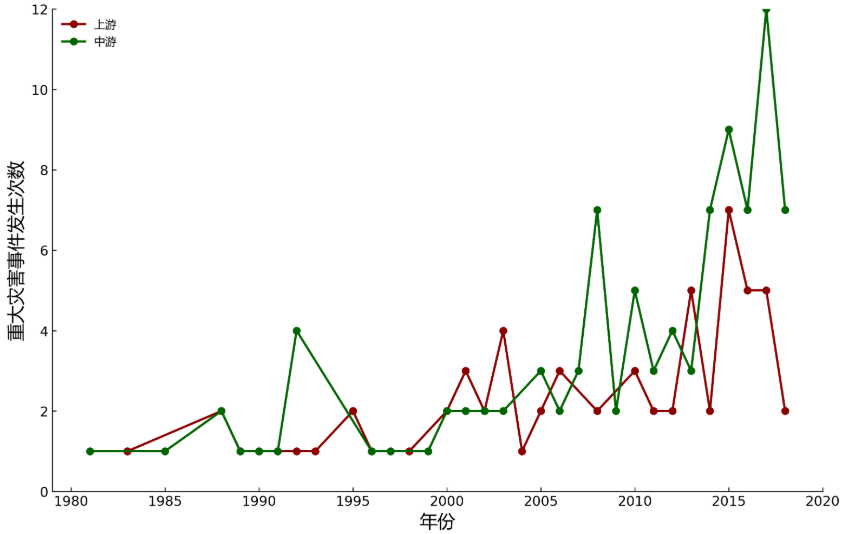


图8 1981—2018年黄河流域重大灾害事件发生次数变化

Fig.8 Changes in the Frequency of Major Disaster Events in the Yellow River Basin From 1981 to 2018

3.2 并发灾害识别与损失评估

并发灾害是指多种灾害在时间与空间上相互重叠的现象，与孤立灾害相比，其影响程度更为显著[25]。由于两个灾害事件的位置与时间不可能完全相同，因此本研究通过在灾害之间设置不同的时间与空间容忍度进行匹配。具体而言，设定重大灾害事件空间影响范围是省（区）级，一般灾害事件空间影响范围是市县级。如果是洪水、暴雨等突发性事件，其灾害持续时间设定为15 d，重大灾害时间设定为1 a（GDIS、EM-DAT等数据集灾害事件的时间粒度是1 a），如果是干旱等持续性事件，其时间设定为3个月。将所有灾害事件根据省（区）市县不同层级地理单元进行分组，如果在同一个区域同一时间段识别到2个以上灾害时间，则认为发生了并发灾害。1992—2019年总共识别到并发灾害37次，具体如图9所示。2017年黄河流域4个省（区）发生了并发灾害，陕西省发生10次，山西省发生4次，甘肃省发生3次，宁夏回族自治区发生2次。

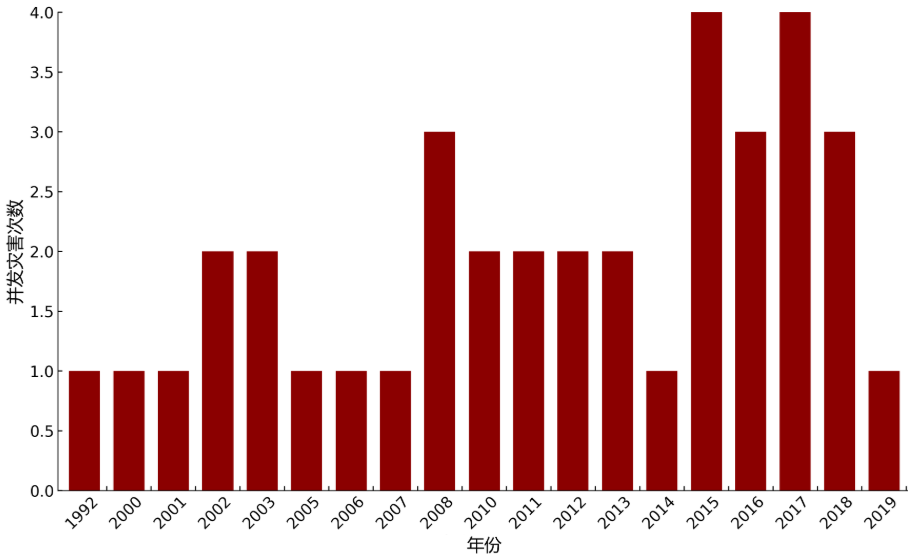


图9 黄河流域历年并发灾害次数

Fig.9 Number of Concurrent Disasters in the Yellow River Basin Over the Years

5 结束语

本文描述了面向灾害损失评估的黄河流域自然灾害知识图谱构建及应用过程。在模式层定义了自然灾害事件本体模型、基础地理信息本体模型与灾害影响本体模型，在多源异构的自然灾害事件数据、基础地理信息数据与灾害损失数据基础上，通过实体与关系识别、知识融合、知识存储构建黄河流域自然灾害知识图谱，同时通过PostgreSQL与Neo4j进行存储和管理，支持时空关系的知识检索和快速识别评估区域并发灾害。

笔者构建的黄河流域自然灾害知识图谱，通过多源异构数据融合与时空关系建模，提升了灾害损失评估的效率，同时突破传统单灾种知识体系的局限，将黄河流域海量复杂的自然灾害信息提炼成知识，对于促进自然灾害领域数据的充分利用与快速分析，提升黄河流域科学决策能力具有一定的理论与应用价值。但随着各类传感器与遥感等新技术的快速发展与应用，如何高效存储与管理海量数据，以及如何无缝接入实时数据流，已成为亟待解决的关键问题。在后续研究中，需要继续完善知识图谱模式层本体建设，扩充数据层来源，从而提高知识图谱的完整度。

**参考文献：**

[1] 兰恒星,彭建兵,祝艳波,等.黄河流域地质地表过程与重大灾害效应研究与展望[J].中国科学:地球科学,2022,52(2):199-221.

[3] 刘峤,李杨,段宏,等.知识图谱构建技术综述[J].计算机研究与发展,2016,53(3):582-600.

[4] 张军珲,黄希扬,桂明宇,等.面向数字孪生工程的水利知识图谱构建及应用[J].人民黄河,2024,46(4):121-124,130.

[5] 刘烨宸,李华昱.领域知识图谱研究综述[J].计算机系统应用,2020,29(6):1-12.

[6] 曾晓玲,张弓.基于黄河资源整合共享的知识图谱研究和应用[J].人民黄河,2021,43(增刊2):282-284.

[7] 朱浩,赵红莉,段浩,等.水利多领域知识图谱关联融合方法研究[J].人民黄河,2025,47(3):123-129,134.

[8] 张军珲,黄希扬,桂明宇,等.面向数字孪生工程的水利知识图谱构建及应用[J].人民黄河,2024,46(4):121-124,130.

[9] 侯梦薇,卫荣,陆亮,等.知识图谱研究综述及其在医疗领域的应用[J].计算机研究与发展,2018,55(12):2587-2599.

[10] 蒲天骄,谈元鹏,彭国政,等.电力领域知识图谱的构建与应用[J].电网技术,2021,45(6):2080-2091.

[11] 侯琛,牛培宇.农业知识图谱技术研究现状与展望[J].农业机械学报,2024,55(6):1-17.

[12] 邱芹军,吴亮,马凯,等.面向灾害应急响应的地质灾害链知识图谱构建方法[J].地球科学,2023,48(5):1875-1891.

[13] 沈伟豪,钟燕飞,王俊珏,等.多模态数据的洪涝灾害知识图谱构建与应用[J].武汉大学学报(信息科学版),2023,48(12):2009-2018.

[14] 王益鹏,张雪英,党玉龙,等.顾及时空过程的台风灾害事件知识图谱表示方法[J].地球信息科学学报,2023,25(6):1228-1239.

[15] 习近平.在黄河流域生态保护和高质量发展座谈会上的讲话[J].中国水利,2019(20):1-3.

[16]兰恒星,祝艳波,李郎平,等.黄河流域地质—地貌—气候多过程相互作用及其孕灾机制研究[J].中国科学基金,2021,35(4):510-519.

[17] 祝艳波,兰恒星,彭建兵,等.黄河中游地区水土灾害机理与灾害链效应研究进展[J].人民黄河,2021,43(8):108-116,147.

[18] ROSVOLD E L,BUHAUG H.GDIS,a Global Dataset of Geocoded Disaster Locations[J].Scientific Data,2021,8(1):61.

[19] 王毅,杨舒楠,张立生,等.三个全球气象灾害数据库对比及展望[J].气候变化研究进展,2022,18(2):253-260.

[20] 郭红翔,朱文泉.社会经济统计数据空间化研究进展[J].地理学报,2022,77(10):2650-2667.

[21] 吴吉东,何鑫,王菜林,等.自然灾害损失分类及评估研究评述[J].灾害学,2018,33(4):157-163.

[22]唐彦东,张佳丽,于汐,等.灾害间接经济损失评估研究综述[J].自然灾害学报,2023,32(6):1-11.

[23] 吴华意,刘波,李大军,等.空间对象拓扑关系研究综述[J].武汉大学学报(信息科学版),2014,39(11):1269-1276.

[24] 黄河上、中、下游的分界点在哪里？上、中、下游的长度各是多少？[J].河南水利与南水北调,2024,53(3):7.

[25] XU Y L,DAI Q,ZHU J X,et al.Increased Significance of Global Concurrent Hazards From 1981 to 2020[J].Earths Future,2024,12(9):e2024EF004490.

【责任编辑 栗 铭】