

引用格式:覃炆扬,郭俊,刘懿,等.数字孪生流域知识图谱构建及其应用[J].水利水电快报,2023,44(11):115-120.

# 数字孪生流域知识图谱构建及其应用

覃炆扬,郭俊,刘懿,舒海润

(华中科技大学 土木与水利工程学院 湖北 武汉 430074)

**摘要:**为通过数字孪生构建物理流域与数字流域关联,实现对流域资源、环境、生态等多方面的监测、分析与优化配置,探讨了数字孪生流域建设过程中知识图谱的应用前景、构建方法、应用和面临的挑战。在知识工程相关技术基础上,以甘肃省白龙江干流舟曲县河段流域为例,对其中水利对象、防洪保护对象及其之间的关联关系进行了知识表示、抽取、融合和推理,构建了可视化、交互式的流域知识拓扑图,使其为数字孪生流域管理提供需求分析、知识检索、决策分析等应用价值,并对可能面临的挑战进行了分析,如需提升数据质量与标准化问题、或将面临的技术瓶颈等。研究成果可为水利工程知识图谱建设与推广应用提供参考。

**关键词:**知识图谱;数字孪生流域;平台构建;白龙江

中图分类号:TP391

文献标志码:A

DOI: 10.15974/j.cnki.slsdkb.2023.11.020

文章编号:1006-0081(2023)11-0115-06

## 0 引言

随着社会经济的发展,水资源、水环境、水生态等问题日益突出。国家“十四五”规划中,明确提出要推进智慧水利体系构建<sup>[1]</sup>,加强水资源保护与利用,加大水污染防治力度,推进流域综合治理,提高防洪排涝能力,保障水安全。为实现这一目标,数字孪生流域作为一种新型的基础设施和流域管理方式应运而生。数字孪生流域是以物理流域为单元、时空数据为底座、数学模型为核心、水利知识为驱动,对物理流域全要素和水利治理管理活动全过程的数字化映射与智能化模拟,可实现与物理流域同步仿真运行、虚实交互与迭代优化<sup>[2]</sup>。数字孪生流域通过构建物理流域与数字流域的关联,实现对流域资源、环境、生态等方面的监测、分析与优化配置。

数字孪生流域建设技术大纲中,包含了数字孪生平台与信息化基础设施。数字孪生平台中,知识平台集成信息来自数据底板的相关数据和模型平台的计算分析结果,这些信息经过水利知识引擎的处理后,形成知识图谱(Knowledge Graph, KG),用来支

撑水利业务的应用<sup>[3]</sup>。知识图谱作为一种新兴的知识表示和管理技术,能够有效地组织、整合和挖掘领域知识,支持构建数字孪生流域的知识平台,为数字孪生流域提供决策支持。

以图谱形式呈现的可用知识库内容即为知识图谱,知识图谱近年来备受关注,并产生了大量的研究,大多数研究都集中在知识图谱的生成及其内部信息的消费<sup>[4-6]</sup>。知识图谱的发展经历了3个阶段:在第一阶段,知识表征被提升到Web标准的水平。在第二阶段,核心关注点转向数据管理、链接数据及其应用。在第三阶段,焦点转向实际应用(包括语义解析<sup>[7-8]</sup>、信息抽取<sup>[9]</sup>、推荐系统<sup>[10]</sup>、问答系统<sup>[11-12]</sup>等)。冯钧等<sup>[12]</sup>总结了领域知识图谱的研究现状,并指出了水利领域知识图谱的研究方向。知识图谱的研究总结工作基本涉及知识抽取、知识存储、知识融合、知识推理、知识表示,水利领域也是在此基础上进行细化研究与应用展示。本文在现有体系基础上,探索性提出一种以流域拓扑图的形式构建与应用的知识图谱,以更好地梳理流域相关的知识脉络,为相关防洪等工作提供更好的支撑。

收稿日期:2023-04-24

基金项目:“十四五”国家重点研发计划“流域性大洪水场景推演及预案关键技术”(2022YFC3002704);湖北省自然科学基金“基于区块链的湖北省小流域防洪数据治理技术研究”(2022CFB042);中国工程院院地合作重点项目“绿水武汉”水利工程数字孪生体系构建与战略研究(HB2022B19)

作者简介:覃炆扬,男,硕士研究生,研究方向为水信息科学与数字流域工程。E-mail: qinyi1207@foxmail.com

通信作者:刘懿,男,副教授,博士生导师,研究方向为水旱灾害防治、基层防汛预报预警体系、水库发电及防洪调度、径流预报、流域水资源管理等方面。E-mail: Prof\_liuyi@hust.edu.cn

### 1 技术路线探究

知识图谱大体上分为自顶向下和自底向上两种构建方式,图 1 为自底向上的技术路线图,揭示了一个迭代更新的知识图谱构建过程。一轮迭代主要分为知识抽取、知识融合和知识加工 3 个阶段。知识抽取也是信息抽取,首先对结构化、半结构化、非结构化数据进行抽取,然后从各数据源中进行实体、属性和关系的抽取,形成本体化的知识表达。从知识抽取、数据库、知识库获取新知识后,需要对知识进行整合,比如同一个实体在多条知识中有不同表示、同一个称谓关联着多个实体等,因此需要进行知识融合,消除歧义与矛盾。融合过的知识,经过本体抽取、质量复核(需要人工参与评估、甄选)后,可用的部分被补充到知识库中,保证知识库的质量维护和更新,知识库中的内容也需定期检查、推理并重新评估。

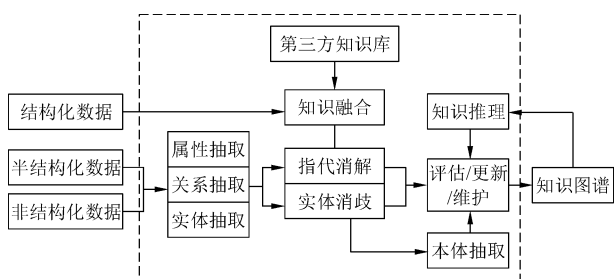


图 1 知识图谱构建路线

Fig. 1 Knowledge graph construction route

### 2 数字孪生流域知识图谱构建

为实现数字孪生流域的智能管理,需要构建一套知识图谱体系架构,涵盖知识表示、知识抽取、知识融合和知识推理等关键技术环节。

#### 2.1 知识表示

知识表示是知识图谱的基础,需要构建一套适用于水利领域的实体、属性和关系表示方法。首先,通过咨询领域专家和查询文献资料,梳理水利领域的基础本体和业务本体,包括江河湖泊、水利工程、水资源管理等实体以及它们之间的关系。其次,采用资源描述框架(Resource Description Framework, RDF)和网络本体语言(Web Ontology Language, OWL),对实体、属性和关系进行结构化的描述,构建水利领域的知识表示体系。将知识图谱看成是由互相连接的实体和属性构成的网络,则每一条知识表示为一个三元组,即由主体(Subject)、谓语(Predicate)及宾语(Object)三部分构成。图 2 是 RDF 三

元组以 RDF/XML 文档形式进行存储的数据格式。

```
<RDF>
  <Description about="http://www.kg_link_1.com/RDF">
    <author>Hi-Brain</author>
    <homepage>http://www.kg_link_1.com/homepage</homepage>
  </Description>
</RDF>
```

图 2 RDF 三元组格式

Fig. 2 RDF triple format

#### 2.2 知识抽取

知识抽取是从原始的结构化、半结构化和非结构化数据中提取结构化知识的过程,涵盖了自然语言处理、机器学习等技术在其中的应用,图 3 展示了各种原始数据的一般分类。在数字孪生流域建设中,数据来源主要包括文本、图像、视频等多模态数据,利用自然语言处理、机器学习和深度学习等技术,对文本、表格和图像等多源异构数据进行有效提取与处理。例如,可以通过命名实体识别、关系抽取等技术,从文本中提取水资源、水环境等相关实体及其属性和关系;通过图像识别和分割技术,从遥感影像中提取水体、地表覆盖等空间信息。通过实体识别、关系抽取和属性抽取等技术,从数据中构建实体-关系三元组,形成水利领域知识图谱的信息基础。



图 3 原始数据类型

Fig. 3 Original data type

#### 2.3 知识融合

知识融合包含实体对齐、关系对齐与知识补全方法。由于多源数据的异构性和不一致性,需要采用知识融合技术对知识进行整合。实体对齐是将描述同一现实世界对象的不同实体进行关联的过程,可以通过基于特征的相似度计算和基于规则的方法进行实现;关系对齐是指将不同来源的相同关系进行关联,可以通过关系匹配和链接技术实现;知识补全是指出通过挖掘已有知识图谱中的潜在规律,补全缺失的实体属性或关系,可以通过基于规则推理、矩阵分解、表示学习、迁移学习和协同过滤等方法进行。知识融合的过程需要充分考虑不同数据源的可信度、时效性和一致性问题,以保证融合后知识的

质量。

## 2.4 知识推理

知识推理是一种通过已有知识推导出新的知识或尚未标注收录知识的过程,可以帮助挖掘实体之间的潜在关系,化简解决复杂问题。知识推理主要有基于描述逻辑的推理引擎和基于规则的推理方法。基于描述逻辑的推理引擎主要利用本体语言中的概念、属性和关系等描述逻辑元素,实现对知识图谱的推理和验证;基于规则的推理方法则通过编写领域专家的经验规则,利用事实和规则进行推理,得出结论。通过知识推理,可以发现隐含的知识,辅助决策者更好地理解流域管理问题,提高决策效率。

## 3 基于知识图谱的数字孪生流域管理

### 3.1 需求分析

需求分析是了解和挖掘流域管理问题的核心目标和关键举措,为知识图谱的应用提供指导。通过与领域专家沟通、参考文献资料分析和政策法规查询,充分考虑流域管理的多目标性、多尺度性和多时空性等特点,确保知识图谱的实用性和针对性,明确数字孪生流域管理的需求,如水资源优化配置、水环境监测与保护、防洪排涝与灾害防治等。

### 3.2 知识检索

知识图谱可以帮助用户快速准确地查找到所需信息。基于语义的检索方法可以实现对实体、属性和关系等知识元素的精确查找,支持复杂查询条件的组合。通过构建水利领域的本体模型,将关键词映射到知识图谱中的实体和关系,以实现基于语义的检索。同时,引入自然语言处理等技术,进一步提高检索的易用性;引入应用推荐系统,为用户提供个性化推荐,提高检索效率。

### 3.3 知识分析

知识分析是对知识图谱中的数据进行挖掘和分析的过程,以发现潜在的规律和关系。基于知识图谱的数据挖掘方法包括关联规则挖掘、聚类分析、分类预测等。例如,通过关联规则挖掘可以发现水资源与水环境、水利工程之间的关联关系;通过聚类分析可以对流域内的水资源、水环境等相关要素指标进行分区划分;通过分类预测可以预测水资源、水环境相关要素指标的未来变化趋势。这些分析方法可以为流域管理决策提供数据支持。

## 3.4 决策支持

基于知识图谱的决策支持系统可以为流域管理决策者提供有针对性的建议。通过知识推理和知识分析的结果,为决策者提供清晰的信息脉络和可视化展示,帮助他们更好地理解问题,专注于制定合适的决策。同时,可以结合专家经验和历史案例,为决策者提供更具参考价值的建议。

## 4 搭载知识图谱的平台构建方法

### 4.1 基于图数据库和三元组的存储方法

知识图谱的存储需要解决大规模、高并发、多模态数据的存储问题。图数据库作为一种专门用于存储图结构数据的数据库,具有良好的扩展性、高效性和易用性等特点,适合作为知识图谱的存储方式。在数字孪生流域建设过程中使用知识图谱存储,可以选择 Neo4j、JanusGraph 的图数据库作为存储系统。图 4 是结合 Neo4j 的图数据库构建示例。



图 4 结合 Neo4j 的图数据库应用示例

Fig. 4 Graph database application example with Neo4j integration

### 4.2 基于分布式的计算方法

知识图谱的管理需要解决数据的处理、计算和分析等问题。基于分布式计算的知识图谱管理方法可以充分利用计算资源,实现大规模知识图谱的高效处理,如 Hadoop、Spark、Flink 等分布式计算框架进行计算管理。

### 4.3 基于 Web 服务的访问方法

知识图谱的访问需要提供统一的接口,以方便

外部应用的调用。基于 Web 服务的知识图谱访问方法可以实现跨平台、跨语言的数据访问,提高知识图谱的互操作性。在数字孪生流域领域,可以构建如 RESTful API 的知识图谱访问服务,实现对知识图谱的查询、修改和删除等操作。

#### 4.4 基于版本控制的维护方法

知识图谱的维护需要解决数据的更新、修订和版本管理等问题。基于版本控制的知识图谱维护方法可以追踪知识图谱的修改历史,保证数据的一致性和完整性。在数字孪生流域建设过程中,可以借鉴 Git 版本控制系统的思想,实现对知识图谱的版本管理。

### 5 白龙江流域知识拓扑图平台应用

平台硬件结构方面,采用 B/S 架构,服务和数据库部署在远端服务器,通过 Web 前端页面进行 UI 访问或通过 API 进行服务调用<sup>[13]</sup>。平台软件设计方面,分为以下 3 层:① 表示层。基于 Web 前端 GoJS(JavaScript 和 TypeScript 库)进行设计,旨在构建交互式图形界面,提升用户平台操作体验,直观形象展现内容。② 业务层。包括读取水利对象信息并进行拓扑图的绘制、查询对象属性和关联信息、修改编辑等功能和提供拓展的应用接口等。③ 数据层。存储水利对象的属性数据和关系结构数据,并保证数据质量、查询效率以及拓展性能。

白龙江位于甘肃省舟曲县的西北部尕瓦山处流入舟曲县境,经流先往南方向,而后流经曲瓦镇,之后转向东南,经巴藏镇、立节镇、憨班镇、峰迭镇、江盘镇和舟曲县城关镇后,经南峪镇、大川乡流入宕昌县。图 5 为平台应用中的流域拓扑图模块,以拓扑图的形式对白龙江干流舟曲县河段进行了知识脉络展示。拓扑图分为节点和链接,分别代表着流域中水利对象和相关关系的映射,从白龙江上游流域为一个节点起,以水流上下游关系和汇流关系作为链接线,连接起代古寺水电站到石门坪水电站之间的所有水利对象(各水电站)、防洪保护对象(各村庄城镇)、各小支流流域和相关河道关键断面等,其中的每一个节点与每一条链接都存储相关的索引,与数据层的对应数据进行绑定(比如地理位置、行政区划、对象属性等),可以通过点击进行数据查询、切换显示内容、信息编辑等操作。该拓扑图的可拓展性强,能与知识图谱动态对应,也能对节点、连线及其属性进行自定义。

通过知识拓扑图,能快速掌握白龙江干流舟曲

县河段流域内的空间联系和水力水文联系,更好地实行防洪措施、水资源调配和水利工程调度。

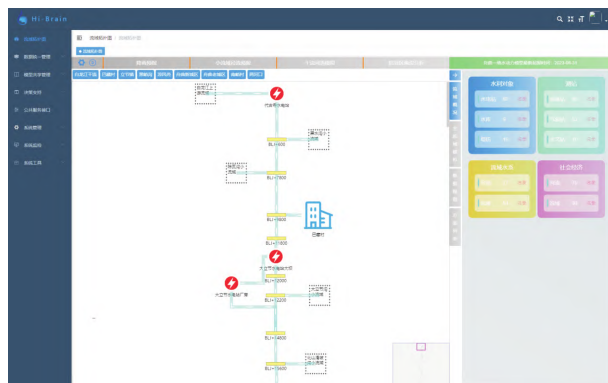


图 5 应用平台上的白龙江流域知识拓扑图

Fig. 5 Knowledge topology of Bailong River Basin on the application platform

### 6 结语

本文介绍了数字孪生流域的概念和知识图谱的相关研究与构建方法,以甘肃省白龙江干流舟曲县河段流域为例,在上游代古寺水电站到下游石门坪水电站的 78 km 沿河区间内,对其中的多个水电站、河道关键断面、划分子流域、防洪保护对象(城集镇)等进行抽象概化。以拓扑图中节点和链接分别代表对象(主、宾)和关系(谓)的形式,对这些水利对象及其之间的关联关系(如上下游关系、汇流关系、所属关系等)进行了知识提取和梳理,映射成可视化、交互式的流域知识拓扑图。

流域知识拓扑图探索了知识图谱在数字孪生流域的应用新形式,其以河流为脉络,将物理流域概化映射到数字平台上,直观地展示流域内的各类水利对象和相关的空间、水力、管理、行政等联系,并且能对绑定的属性等数据进行查询、修改,具有部署灵活、方便用户访问等优势,可在水资源管理、防洪、数字沙盘等方面提供支持。

### 7 展望

数字孪生流域建设的新兴性,以及它在结合知识图谱应用方面所展现出的创新价值,揭示了未来的研究方向应进一步拓宽并面临着多方面的挑战。为推动数字孪生流域和知识图谱技术深度融合与持续发展,提出以下几点展望。

(1) 提升数据规模与质量。数字孪生流域未来将涉及大量不同类型的数据,并且面临数据异构、不一致和不完整等问题。随着数据的不断积累和领域知识的更新,知识图谱需要不断地进行动态维护和

优化。如何实现知识图谱的自动化构建与更新,确保其时效性和可靠性,以适应流域管理的实时性需求,是亟待解决的问题。

(2) 进行跨学科领域的知识融合与交流。数字孪生流域涉及水文学、气象学、地理信息科学等多个学科领域。如何在知识图谱中实现跨领域知识的融合以及促进不同领域专家之间的交流与合作,是知识平台建设中需要克服的难题。

(3) 数据安全与保护。如何在知识库构建与应用中平衡数据共享与数据保护的需求,以及如何确保知识图谱中的数据和知识的合规性和伦理性,是在实际应用过程中需要关注的问题。

参考文献:

- [1] 习近平总书记关于《中共中央关于制定国民经济和社会发展第十四个五年规划和二〇三五年远景目标的建议》的说明[J]. 系统工程 2022 40(5):159.
- [2] 水利部部署数字孪生流域建设工作[J]. 中国水利, 2022(1):5.
- [3] 谢文君,李家欢,李鑫雨,等.《数字孪生流域建设技术大纲(试行)》解析[J]. 水利信息化 2022(4):6-12.
- [4] SUCHANEK F M ,KASNECI G ,WEIKUM G. Yago: a core of semantic knowledge [C]//Proceedings of the 16th international conference on World Wide Web. New York: Association for Computing Machinery 2007.
- [5] BOLLACKER K ,EVANS C ,PARITOSH P ,et al. Freebase: a collaboratively created graph database for structuring human knowledge [C] // Proceedings of the 2008 ACM SIGMOD international conference on Management of data. New York: Association for Computing Machinery , 2008.
- [6] VRANDEČIĆ D ,KRÖTZSCH M. Wikidata: a free collaborative knowledgebase [J]. Communications of the ACM , 2014 57( 10) : 78 - 85.
- [7] BERANT J ,CHOU A ,FROSTIG R ,et al. Semantic parsing on freebase from question - answer pairs [C]//Proceedings of the 2013 conference on empirical methods in natural language processing. Stroudsburg: Association for Computational Linguistics 2013.
- [8] HECK L ,HUANG H. Deep learning of knowledge graph embeddings for semantic parsing of twitter dialogs [C]//2014 IEEE Global Conference on Signal and Information Processing ( GlobalSIP) . Piscataway: IEEE 2014.
- [9] DIETZ L ,XIONG C ,DALTON J ,et al. Special issue on knowledge graphs and semantics in text analysis and retrieval [J]. Information Retrieval Journal 2019 22: 229 - 231.
- [10] SUN R ,CAO X ,ZHAO Y ,et al. Multi - modal knowledge graphs for recommender systems [C]//Proceedings of the 29th ACM international conference on information & knowledge management. New York: Association for Computing Machinery 2020.
- [11] SAXENA A ,TRIPATHI A ,TALUKDAR P. Improving multi - hop question answering over knowledge graphs using knowledge base embeddings [C]//Proceedings of the 58th annual meeting of the association for computational linguistics. Stroudsburg: Association for Computational Linguistics 2020.
- [12] 冯钧,杭婷婷,陈菊,等.领域知识图谱研究进展及其在水利领域的应用[J]. 河海大学学报(自然科学版) 2021 49(1):26-34.
- [13] 饶小康,马瑞,张力,等.数字孪生驱动的智慧流域平台研究与设计[J]. 水利水电快报 2022 43(2):117-123.

(编辑:唐湘茜 张 爽)

## Digital twin watershed knowledge graph construction and its application

QIN Yangyang ,GUO Jun ,LIU Yi ,SHU Hairun

( School of Civil and Hydraulic Engineering ,Huazhong University of Science and Technology ,Wuhan 430074 ,China)

**Abstract:** In order to construct the association between physical watershed and digital watershed through digital twin and realize the monitoring ,analysis and optimal allocation of resources ,environment ,ecology and other aspects of the watershed ,the application prospect ,construction method ,application and challenges faced by the knowledge map during the construction of digital twin watershed were discussed. On the basis of knowledge engineering related technology ,taking the river basin of Zhouqu County section of the main stream of Bailong River in Gansu Province as an example ,the knowledge representation ,extraction ,fusion and reasoning of water conservation objects ,flood protection objects and their associated relationships were carried out ,and the visualized and interactive knowledge

topology map of the basin was constructed. The value of demand analysis ,knowledge retrieval ,decision – making a– nalysis and other applications for the management of digital twin river basin were provided. The possible challenges were analyzed ,such as the need to improve data quality and standardization or will face technical bottlenecks ,etc. The research results can provide a reference for the construction and promotion of the application of knowledge map of water conservancy engineering.

**Key words:** knowledge graph; digital twin watershed; platform construction; Bailong River

~~~~~  
(上接第 114 页)

## Community structure of hydrobiology and water ecological assessment of Dongting Lake in 2020

PENG Xian<sup>1</sup> ,ZHU Dandan<sup>1</sup> ,WANG Chouming<sup>2</sup>

(1. Yueyang Ecology Environmental Monitoring Center of Hunan ,Yueyang 414000 ,China; 2. Ecological and Environmental Mo– nitoring Center of Dongting Lake ,Yueyang 414000 ,China)

**Abstract:** To grasp the community structure of hydrobiology and assess water ecological of the Dongting Lake ,two surveys of hydrobiology in Dongting Lake were carried out in 2020. The comprehensive index of water ecological en– vironmental quality was used to evaluate the water ecology. The results showed that a total of 6 phyla and 73 genera of phytoplankton were detected in Dongting Lake ,the dominant species were *Pseudanabaena sp.* ,*Melosira sp.* , *Raphidiopsis sp.* ,*Merismopedia sp.* and *Crucigenia sp.* . The average density of phytoplankton was  $95.7 \times 10^4$  /L. A total of 48 genera of zooplankton were detected in Dongting Lake ,the dominant species were *Polyarthra vulgaris* , *Keratella cochlearis* ,*Bosmina sp.* and *nauplius*. The average density of zooplankton was 74 /L. A total of 57 species of benthos were detected ,the dominant species were *Cryptochironomus rostratus* ,*Polypedilum sp.* ,*Bellamyia aerugino– sa* ,*Corbicula fluminea* ,*Branchiura sowerbyi* and *Grandidierella sp.* . The average density of benthos was 141 /m<sup>2</sup>. The overall water quality of Dongting Lake was lightly polluted ,the main item exceeding the standard was the total phosphorus. The comprehensive nutritional status index was 46. 87 ,indicating the moderate nutrition. The evaluation result of habitat quality in Dongting Lake was 92 ,indicating the medium nutrition. WQI index was used to evaluate the water ecology of Dongting Lake ,the results showed that the water ecological environment quality of Dongting Lake was good.

**Key words:** hydrobiology; community structure; water quality assessment; water ecological assessment; Dongting Lake



欢迎关注“长江水利”微信公众号