

# 河道健康管理本体模型构建与数据存储方法设计

刘晓艳<sup>1,2,3</sup>, 田兆炜<sup>1,2</sup>, 周静怡<sup>1,2</sup>, 赵天浩<sup>1,2</sup>, 徐颖<sup>4</sup>, 徐家鹏<sup>5</sup>, 沈婕<sup>1,2,3</sup>

(1. 南京师范大学虚拟地理环境教育部重点实验室, 江苏南京 210023; 2. 南京师范大学地理科学学院, 江苏南京 210023;

3. 江苏省地理信息资源开发与利用协同创新中心, 江苏南京 210023;

4. 山东东方道途数字数据技术有限公司, 山东济南 250000; 5. 北京百度智图科技有限公司, 北京 100193)

**摘要:** 河道的有效管理是维持河道健康生态状况的前提条件。针对河道健康数据存在的跨部门管理、存储效率低下、语义信息被忽略等问题, 提出了一种基于本体的河道健康数据存储方法。构建了河道健康领域的本体模型, 设计了基于 Neo4j 图数据库的河道健康数据存储方法。以上海市宝山区河道为例, 开发了河道健康可视化原型系统, 实现了河道健康本体数据在概念、关系、属性和实例的存储与查询。设计了 2 种不同本体存储方式查询效率的对比试验, 验证了本文方法的有效性和可行性。

**关键词:** 河道健康; 河道管理; 本体; 图数据库; 数据存储

中图分类号: P208; X321

文献标志码: A

health data based on Neo4j graph database is designed. Taking the river in Baoshan District, Shanghai as an example, a river health visualization prototype system is developed, which successfully achieves the storage and query of river health ontology data in concepts, relations, attributes and instances. A comparison test of the query efficiency of two different ontology storage methods is designed to validate the effectiveness and feasibility of the method proposed in this paper.

**Key words:** river health; waterway management; ontology; graph database; data storage

## Ontology Model Construction and Data Storage Method Design for River Health Management

LIU Xiaoyan<sup>1,2,3</sup>, TIAN Zhaowei<sup>1,2</sup>, ZHOU Jingyi<sup>1,2</sup>, ZHAO Tianhao<sup>1,2</sup>, XU Ying<sup>4</sup>, XU Jiapeng<sup>5</sup>, SHEN Jie<sup>1,2,3</sup>

(1. Key Laboratory of Virtual Geographic Environment of the Ministry of Education, Nanjing Normal University, Nanjing 210023, China; 2. School of Geography, Nanjing Normal University, Nanjing 210023, China; 3. Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China; 4. Shandong Eastdawn Co., Ltd., Jinan 250000, China; 5. Beijing Baidu Zhitu Technology Co., Ltd., Beijing 100193, China)

**Abstract:** Effective management of river channels is a prerequisite for maintaining the healthy ecological condition of river channels. An ontology-based method for storing multi-source river health data is proposed to address the problems of cross-sector management, inefficient storage, and semantic information ignored in river health data. An ontology model of river health domain is constructed, and a storage method of river

健康的水环境是人类赖以生存的根本条件, 是保证社会可持续发展的重要基础, 也是实现生态文明、建设美丽中国的重要保障。2015 年联合国提出《可持续发展目标》, 强调为所有人提供水和环境卫生并对其进行可持续管理<sup>[1]</sup>。近年来, 随着移动技术、物联网、云计算技术的发展与普及, 我国水务信息化建设逐步深入, 有力地提高了智慧水务、智慧城市的建设水平<sup>[2]</sup>。其中, 与人类生活息息相关的健康的河道生态系统日益受到重视。河道的有效管理是维持河道健康生态状况的前提条件, 为了加强河道管理, 测绘、环境、水务、规划等部门进行协同管理, 利用各种监测或统计手段, 获取城市区划、河道监测、河道相关人员信息等大量数据。传感器多时段、全天候的数据获取能力使河道自动监测的数据种类更加丰富、精度更加准确、时效性更为突出。河道管理人员通过定期巡查, 实现对河道水质情况、排污口等信息记录并上报。丰富的数据获取手段为河道健康的评估与应用提供了优质的数据支持。但是对于河道健康数据的处理与应用往往局限于本部

收稿日期: 2023-05-03

基金项目: 国家重点研发计划(2021YFE0112300)

第一作者: 刘晓艳(1974—), 女, 讲师, 理学博士, 主要研究方向为环境制图。E-mail: liuxiaoyan@njnu.edu.cn

通信作者: 周静怡(1998—), 女, 博士生, 主要研究方向为环境制图、室内导航地图。E-mail: zhoujingyi@njnu.edu.cn



论文  
拓展  
介绍

门,并存在存储效率较低、语义信息被忽略等问题,面向河道健康领域数据的集成与存储方法需要进一步研究。

本体在多源异构数据语义异构处理上具有明显优势,国内外学者在地理学、环境学、生态学等领域做了大量本体构建研究。如Lee等<sup>[3]</sup>从非结构化文本中抽取知识,并基于事件构建本体;Hassanein等<sup>[4]</sup>提出了一个尼罗河地理本体模型,并基于该本体实现上下文的智能搜索查询;王蔚华等<sup>[5]</sup>通过解析生态水文的内涵,以黑河流域为研究区域,提出了基于知识的生态-水文本体的构建方法。在多源数据集成与存储方面,崔巍等<sup>[6]</sup>研究了一种基于本体的GIS集成结构,从而提升了地理信息系统间在语义上的互操作能力;Bellini等<sup>[7]</sup>为解决智能城市建设中数据来源众多但共享困难的问题而构建智能城市本体;对于本体数据的存储主要基于图数据库实现,如宫法明等<sup>[8]</sup>研究了油田本体,提出了一种基于Neo4j(是一个高性能的NoSQL图形数据库,其数据存储结构和数据查询方式都是以图论为基础)的领域本体构建方法,减少其存储空间和提升查询效率;王红等<sup>[9]</sup>构建了民航突发事件应急管理本体,并提出了一种基于Neo4j的领域本体数据存储方法;Comyn-Wattiau等<sup>[10]</sup>基于模型驱动方法将NoSQL属性图数据库属性拓展到实体关系的模式。

本体已经被学者们应用于不同领域的数据集成与存储研究,但在河道健康领域的应用研究尚不多见。如何结合河道健康数据特点,利用领域本体对河道健康数据进行有效集成存储是值得研究的问题。本文面向河道健康数据集成和管理的实际需求,探讨一种基于本体的河道健康数据存储方法。以上海市宝山区河道为例,收集了丰富的河道健康相关数据,构建了宝山区河道健康本体案例,设计图数据库模式并实现该河道健康管理要素、河道健康评价要素以及河道健康演化过程的存储与查询,验证了本文方法的有效性。

## 1 河道健康本体模型构建

### 1.1 河道健康内涵与场景概念模型

河流健康与人类社会的发展息息相关,随着国内外河流环境的不断恶化,河流健康问题引起了相关领域内学者们的关注。20世纪末,Karr<sup>[11]</sup>在生态完整性以及人类价值观的理论基础上提出河流健康概念。概念一经提出,便迅速成为河流保护与修复

领域的研究热点。因为河流生态系统自身复杂的特点以及学者研究视角的不同,对河流健康的概念有不同的解释,尚未形成统一认识<sup>[12]</sup>。根据研究者关注的重点,河流健康内涵的理解大体可以分为三类:一是从保护生物多样性的角度出发,认为生态系统结构的稳定是河流是否健康的前提条件;二是从人类的利用价值的角度出发,不能仅从生态系统的完整性上判断河流是否健康,还应考虑社会服务功能的重要性<sup>[13]</sup>;三是从河流管理的要求出发探究河流健康的概念与内涵,强调健康的河流就是河流管理达到某种特定的目标<sup>[14]</sup>。本文的研究对象主要为城市河道,城市河道与自然河道的区别在于前者与人类活动的关系更加密切,与社会经济的发展有着密不可分、相互影响与作用的关系。因此,在研究其内涵的时候,既要考虑河流的自然状态和完整的生态系统结构,还要考虑河流的社会属性以及人类对河流的管理需求。

基于上述对河道健康内涵的解析,并根据Studer等<sup>[15]</sup>提出的本体定义,本文对河道健康领域的本体给出了相应的定义,即河道健康本体是“河道健康领域范围内各概念及其关系概念化、形式化的描述与说明”。根据上述概念,参考闫国年等<sup>[16]</sup>提出的地理场景六要素,对河道场景中的人物、事物与事件要素进行归纳,建立了如图1所示的河道场景概念模型。河道场景中的每一个要素都可以通过时间、空间与语义特征的表达来刻画其状态,同时各要素对自身的演化过程进行记录,进而参与到其他要素的演化过程中。

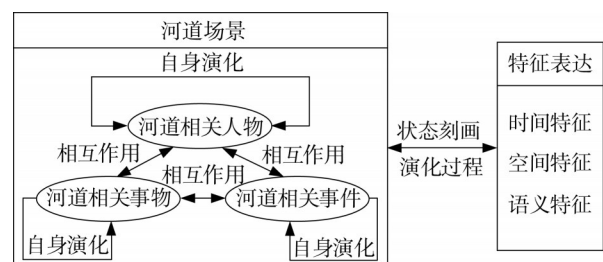


图1 河道场景概念模型

Fig. 1 Conceptual model of river scene

从整体上看,河道健康评价指标体系需要具备明显的层次结构。从单一指标看,指标体系中的每个指标都需要客观反映出河道生态系统的健康状况。结合国内外学者、国家以及相关机构的研究基础和评价指标选取原则,建立城市中小型河道健康评价指标层次结构图(图2),将河道健康指标体系分为不同层次。

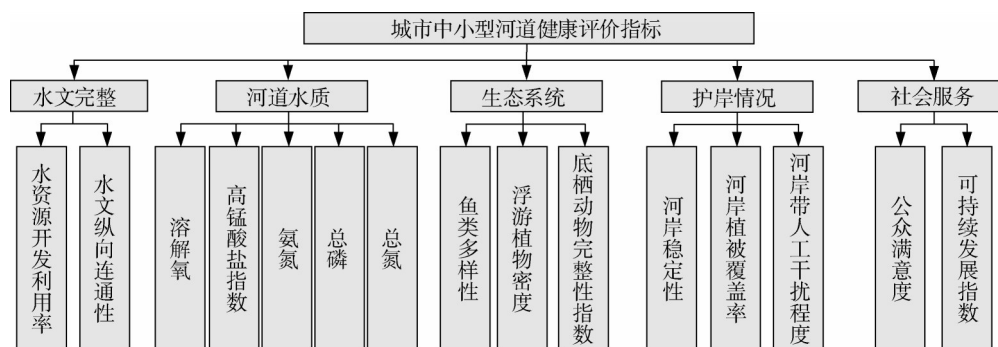


图 2 城市中小型河道健康评价指标层次结构

Fig. 2 Hierarchy of indicators for evaluating the health of urban small and medium-sized rivers

### 1.2 河道健康本体模型

根据河道健康内涵与河道场景概念模型设计河道健康本体概念体系,包括河道相关人物、事物与事

件 3 个一级类。在此基础上,将河道健康本体概念细分为多个层次,具体的河道健康本体概念体系如图 3 所示。

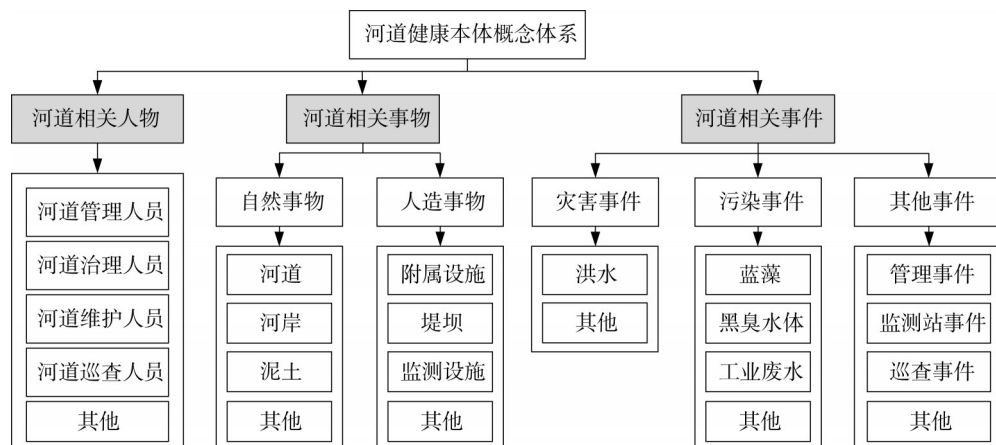


图 3 河道健康本体概念体系

Fig. 3 Ontology concept system of river health

一个完整的河道健康本体主要由概念、关系、属性和实例四部分组成,结合对河道健康内涵的分析,河道健康的本体应考虑 3 个方面:①从河道管理角度出发,需要将河道与管理者、监测站和河道相关事件的关系进行概念化、形象化的描述说明。②从河道评价角度出发,指标是衡量河道是否健康的重要标准,通过河道健康评价指标的描述及其数值展示,让公众更直观地了解河道健康状况。③从河道演化过程出发,对河道及其相关要素的过程、状态与事件在不同时间段的变化进行表达,更好体现出河道这一地理实体的时空特征。

河道健康管理本体是由河道与河道管理人员、监测站和河道相关事件构成。河道作为一个地理实体,本身具有一定的属性特征,会产生相应的事件,同时事件又会驱动河道不断发生变化。河道水质监测数据可以反映相关事件的发生。管理人员作为河

道的管理者,则通过一些措施处理相关事件。

河道健康评价本体根据前文提出的城市中小型河道健康评价指标层次结构构建。河道健康评价包含水文完整、河道水质、生态系统、护岸情况和社会服务 5 个类,它们在语义层面上是兄弟关系。

河道健康演化本体主要由河道健康要素、过程、状态及其之间的关系组成。河道健康要素包括河道要素以及河道管理者要素,两者的演化过程存在一定联系。每个要素都有各自的发展过程,其过程具有零到多个状态。同时,不同状态之间具有一定的时间关系。

## 2 基于 Neo4j 的河道健康数据存储规则设计

通过 Neo4j 图数据库对河道健康本体进行存



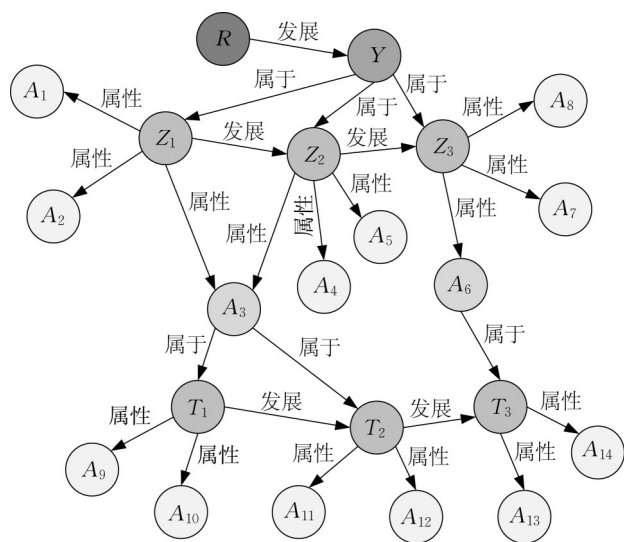


图 7 河道健康演化过程要素存储规则

Fig. 7 Element storage rules of river health evolutionary process

### 3 案例实现

#### 3.1 研究区概况与数据准备

选取上海市宝山区作为研究区域。宝山区地处上海北部,东临黄浦江,由于沿海的地理位置,区内河道、湖泊众多,共有 932 条河道,河道长度为 806.55km,河网密度约为  $2.98\text{km}\cdot\text{km}^{-2}$ 。其中市管河道 10 条,区管河道 28 条,镇管河道 111 条,村级河道 724 条,其他河道和湖泊 59 条。

研究数据来源广泛、类型丰富,主要包括宝山区基础地理信息数据、宝山区地表水原始监测数据、宝山区河道健康评价数据以及宝山区河道健康元数据,如表 1 所示。其中,宝山区地表水原始监测数据来源于上海市公共数据开放平台,通过上海市生态环境局提供的公共数据 API 接口获得;通过检索与整理上海市和宝山区统计局 2019 年统计年鉴,获取宝山区内区级河道的河道健康评价数据。

表 1 研究区数据准备

Tab. 1 Data preparation of study area

数据类型	数据分类	具体内容
基础地理信息数据	河道矢量数据 行政区划数据	宝山区境内市级、区级、镇级以及村级河道编码、等级、流经地区、长度等 宝山区下属区县名称、区县编码等
地表水原始监测数据 河道健康评价数据	地表水河流原始监测数据 2019 年统计年鉴数据	城市代码与名称、河流代码与名称、断面代码与名称、监测年月等 宝山区境内区级河道水质指标
河道健康元数据	河道实体数据 河道管理者实体数据 河道事件实体数据	宝山区各河道相关文件名称、分类以及相关描述 宝山区各河道对应管理者名称、管理者描述、管理者历史行为等 宝山区各河道事件发生时间、地点、事件描述等

#### 3.2 河道健康可视化原型系统

基于上海市宝山区河道健康相关的数据,构建该区域的河道健康可视化原型系统,系统界面如图 8 所示。系统的设计框架是将存储在 Neo4j 中的河道健康数据导出为 json 格式,然后通过 ECharts(提供了一系列直观、可交互、可个性化定制的数据可视化图表,并且可以与各种前端框架和数据处理工具进行集成)调用数据,实现河道与监测站信息的可视化,完成地图与知识图谱的联动显示。通过交互点击地图中的河道或者监测站,系统会以知识图谱的形式展示该河道或者监测站相关信息。

系统主要功能为河道健康地图可视化和知识图谱表达。河道健康地图可视化是指河道健康有关的专题要素在时间与空间上的动态展现,如河道水质情况、河道治理情况等;知识图谱表达实现了河道健康地图中各种要素在语义上的表达,如某条河道或某个监测点的关系展示,本质是对河道健康涉及到

的各种数据的形式化表达,挖掘数据之间的联系,实现数据间的语义关联。

#### 3.3 数据查询性能评价

本体的查询性能是衡量本体数据存储的重要指标。本文以河道健康管理本体为例,设计对比实验,对提出的本体数据存储模式进行验证与评估。设计了 4 类基本的查询问题,对应 5 个具体查询案例,即类查询(Q1:查询 river 类的兄弟类)、关系查询(Q2:查询 river 类与管理者类的关系;Q3:查询 river 类实例与 manager 类实例的关系)、属性查询(Q4:查询 river 类的所有属性)和实例查询(Q5:查询 river 类的所有实例)。其中关系查询中对应了 2 个查询案例,分别为类之间的关系查询与实例之间的关系查询。分别使用 Cypher 语言(一种声明式图数据库查询语言,具有丰富的表现力,能高效地查询和更新图数据)和 SPARQL 语言(全称为 SPARQL Protocol and RDF Query Language,是为 RDF 开发的一种查询语言和数据获取协议,用于任何可以用 RDF 来表



图 8 河道健康地图可视化界面

Fig. 8 Visualization interface of river health map

示的信息资源) 进行查询,为了减少实验结果中偶然因素的影响,对于每一个查询案例都重复执行 20 次,然后取其均值作为最后的实验结果。最终的查询结果如图 9 所示。

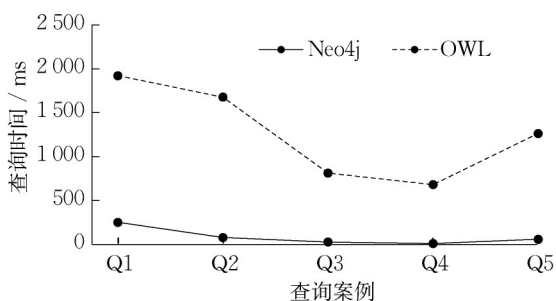


图 9 查询结果对比

Fig. 9 Comparison of query results

经过对 2 种查询方式的效率和稳定性进行分析,基于 Neo4j 存储的本体数据查询消耗时间明显低于基于 OWL (Web Ontology Language, 是一个基于描述逻辑的语言,用于表示和共享复杂的概念和知识结构) 文件本体存储。本文提出的基于 Neo4j 的河道健康数据存储方法在查询效率和稳定性上具备一定的优势。通过建立原型系统与案例验证,本研究实现了河道健康有关数据在时间、空间和语义上的联动展示,为河道健康状况的监测、预警等服务提供支持,有效验证了本文基于本体的河道健康数据存储方法的有效性。

## 4 结论

针对河道健康数据处理与应用中存在的问题,基于本体在多源异构数据的语义异构处理上的优势,提出了面向河道健康数据的存储方法。通过对河道健康内涵的解析,构建了河道场景概念模型,并在此基础上构建了河道健康领域的本体模型;然后建立了河道健康领域本体与图模型之间的映射关系,并设计了基于 Neo4j 图数据库的河道健康数据存储方法,实现了本体向图数据库的转换;最后以上海市宝山区河道为例,搭建了河道健康可视化原型系统,实现了对多源异构的河道健康数据的存储,验证了本文方法的有效性和可行性。

本文在河道健康本体构建及其数据存储方法上取得了一些成果,但需要指出的是,尚未考察河道健康本体推理能力,后续研究将设计相关实验,以更好突出使用本体模型相较于传统数据库的优势;此外,本文对于河道健康本体的构建主要采用人工构建方式,针对大规模数据的存储和集成则应探究更加高效的半自动化和全自动化本体构建方法。

### 作者贡献声明:

- 刘晓艳: 研究总体负责、主要内容撰写。
- 田兆炜: 数据分析、本体模型构建。
- 周静怡: 研究构思、稿件修订与审核。
- 赵天浩: 数据存储规则设计、系统性能评价。
- 徐颖: 数据采集与处理。

徐家鹏:案例研究原型系统实现。

沈 婕:研究方案设计、稿件修订。

### 参考文献:

- [1] 傅伯杰. 联合国可持续发展目标与地理科学的历史任务[J]. 科技导报, 2020, 38(13): 19.  
FU Bojie. UN Sustainable development goals and historical mission of geography [J]. Science & Technology Review, 2020, 38(13): 19.
- [2] 张岩, 张磊. 论智慧水务平台科研数据管理及人工智能技术的应用[J]. 智能建筑与智慧城市, 2020(3): 90. DOI:10.13655/j.cnki.ibci.2020.03.036.  
ZHANG Yan, ZHANG Lei. Discussion on science research data management and application of artificial intelligence technology in smart water platform [J]. Intelligent Building & Smart City, 2020 (3) : 90. DOI: 10.13655/j.cnki.ibci. 2020. 03.036.
- [3] LEE C S, KAO Y F, KUO Y H, *et al.* Automated ontology construction for unstructured text documents [J]. Data & Knowledge Engineering, 2007, 60(3): 547.
- [4] HASSANEIN A M D E, EL HOUBY E M F . Towards a comprehensive Nile River geographical ontology model [J]. Advances in Natural and Applied Sciences, 2012, 6(6): 772.
- [5] 王蔚华, 邹松兵, 肖洪浪, 等. 内陆河流域生态-水文本体的构建方法及应用[J]. 冰川冻土, 2014, 36(5): 1280.  
WANG Weihua, ZOU Songbing, XIAO Honglang, *et al.* Eco-hydrological ontology in inland river basin: Construction method and application [J]. Journal of Glaciology and Geocryology, 2014, 6(5): 1280.
- [6] 崔巍, 蒋天发, 张德新. 用数据挖掘和本体实现空间信息系统语义互操作[J]. 武汉理工大学学报(交通科学与工程版), 2004(1): 118.  
CUI Wei, JIANG Tianfa, ZHANG Dexin. Using data mining and ontology to realize interoperability geographic information systems [J]. Journal of Wuhan University of Technology (Transportation Science & Engineering), 2004(1): 118.
- [7] BELLINI P, BENIGNI M, BILLERO R, *et al.* Km4City ontology building vs data harvesting and cleaning for smart-city services[J]. Journal of Visual Languages & Computing, 2014, 25(6): 827.
- [8] 宫法明, 李翀然. 基于Neo4j的海量石油领域本体数据存储研究[J]. 计算机科学, 2018. 45(S1): 549.  
GONG Faming, LI Xiaoran. Research on ontology data storage of massive oil field based on Neo4j [J]. Computer Science, 2018. 45(S1): 549.
- [9] 王红, 张青青, 蔡伟伟, 等. 基于Neo4j的领域本体存储方法研究[J]. 计算机应用研究, 2017, 34(8): 2404.  
WANG Hong, ZHANG Qingqing, CAI Weiwei, *et al.* Research on storage method of domain ontology based on Neo4j [J]. Application Research of Computers, 2017, 34(8) : 2404.
- [10] COMYN-WATTIAU I, AKOKA J. Model driven reverse engineering of NoSQL property graph databases: The case of Neo4j[C]//2017 IEEE International Conference on Big Data. [s.l.]: IEEE, 2017: 453-458.
- [11] KARR J R. Defining and measuring river health [J]. Freshwater Biology, 1999, 41(2): 221.
- [12] NORRIS R H, THOMS M C. What is river health? [J]. Freshwater Biology, 1999, 41(2): 197.
- [13] JOHNSON L B, BRENNEMAN D H, RICHARDS C. Macroinvertebrate community structure and function associated with large wood in low gradient streams [J]. River Research and Applications, 2003, 19(3): 199.
- [14] ROGERS K, BIGGS H. Integrating indicators, endpoints and value systems in strategic management of the rivers of the Kruger National Park [J]. Freshwater Biology, 1999, 41 (2): 439.
- [15] STUDER R, BENJAMINS V R, FENSEL D. Knowledge engineering: Principles and methods [J]. Data & Knowledge Engineering, 1998, 25(1/2): 161.
- [16] 闫国年, 俞肇元, 袁林旺, 等. 地图学的未来是场景学吗? [J]. 地球信息科学学报, 2018, 20(1): 1.  
LV Guonian, YU Zhaoyuan, YUAN Linwang, *et al.* Is the future of Cartography the scenario science? [J]. Journal of Geo-information Science, 2018, 20(1): 1.