

大数据驱动的世界运河文化资源数据库构建

付宇鹏

山东外国语职业技术大学, 山东 日照 276826

DOI:10.61369/EDTR.2026080009

摘 要 : 运河文化资源的数字化转型已成为世界运河文化资源保护传承和利用的必然趋势。本文探讨大数据在运河文化资源数据库建设中的应用场景, 基于数据收集融合、大数据挖掘及应用、数据服务三部分内容开展研究, 指出运河文化资源数据库建设应开展面向跨类型、跨平台和异构大数据的语义组织、利用机器学习方法的资源内容分类、基于时空演化规律分析的应用技术研究, 进而提高运河文化资源的数据存储与管理效能, 推进文物数字化保护的发展。

关 键 词 : 运河文化; 大数据; 知识图谱; 机器学习; 数字人文

Construction of a Big Data-Driven World Canal Cultural Resource Database

Fu Yupeng

Shandong Foreign Languages Vocational and Technical University, Rizhao, Shandong 276826

Abstract : The digital transformation of canal cultural resources has become an inevitable trend in the protection, inheritance, and utilization of world canal cultural heritage. This paper explores the application scenarios of big data in the construction of canal cultural resource databases, conducting research based on three key components: data collection and integration, big data mining and application, and data services. It highlights that the construction of canal cultural resource databases should focus on semantic organization for cross-type, cross-platform, and heterogeneous big data, resource content classification using machine learning methods, and applied technology research based on spatiotemporal evolutionary pattern analysis. These efforts aim to enhance the efficiency of data storage and management for canal cultural resources and promote the development of digital preservation of cultural relics.

Keywords : canal culture; big data; knowledge graph; machine learning; digital humanities

运河既是水利工程, 同时承载着文明互鉴的历史使命, 从京杭大运河到苏伊士运河, 从巴拿马运河到阿姆斯特丹运河网, 这些贯通古今的运河工程彰显了人类改造自然的智慧, 赋予了运河深厚的文化内涵和历史记忆。传统的文化资源管理模式难以应对海量、异构、多变的运河文化大数据, 亟须利用大数据技术完成运河文化资源的数字化重构, 建设智能化的文化资源数据库, 不仅能够实现运河文化资源的有效集成, 还能够揭示其间蕴含的潜在文化关联, 为运河文化资源的保护传承与创新应用提供科学依据。

一、世界运河文化资源数据采集与整合技术体系

(一) 多源异构运河文化数据采集技术框架设计

运河文化遗产涵盖文献资料、考古发现、遥感影像、社交媒体等多种数据类型, 各具不同的结构特征与信息属性。构建综合性数据采集体系需要统筹多种技术手段, 实现信息的系统化收集。应建立统一的数据采集标准和规范, 减少信息孤岛现象, 促进信息的互联互通; 部署传感器网络实时监测运河遗产状态。通过水质传感器、结构应变传感器等设备持续感知运河环境参数和结构健康状况, 为遗产保护提供数据支撑^[1]。

(二) 基于知识图谱的运河文化资源语义化组织方法

知识图谱技术通过实体、属性、关系构成的三元组形式, 构建运河文化资源的立体化知识网络, 实现诸如古桥等文物与其建造者、使用者、相关历史事件之间的深层语义关联。本体构建过程需要系统梳理核心概念, 建立层次化的概念体系, 如水工设施类别下设船闸、堤坝、码头等子类, 文化活动类别涵盖祭祀、集市、节日等子项。运用自然语言处理技术对历史典籍进行深度挖掘, 自动抽取其中的人名、地名、时间、事件等实体信息, 并识别实体间的关联关系, 将非结构化文本数据转换为结构化的知识图谱。

项目信息: 本文系教育部中外人文交流中心与文化旅游行业中外人文交流研究院2024年度人文交流专项研究课题“基于大数据与人工智能的世界运河文化数字化保护与智能传播平台构建”(项目编号: CCIPEWHLY2024007)科研成果。

（三）跨区域运河文化数据标准化与质量控制机制

因为不同国家的运河文化数据库管理工作不尽相同，有必要制定数据映射模式来完成系统之间的互操作。制定统一元数据标准，包括资源标识、资源描述、权限控制等，以便于对文化资源实行全面的统一规范，在此过程中可以参考国外标准化规定的制定方案并做本土化的改革处理。制定多方位、多层次数据质量评估机制，根据数据的完备、规范、及时、标准等因素，通过程序自动检查和人类主观评测来检查和分析运河文化数据库资源数据质量中存在的异常问题，及时修正数据。按照一定的操作流程完成数据的整理，包括对数据进行删除、补充、纠正错误、规范化数据等^[2]。

二、大数据分析驱动的运河文化资源挖掘与价值发现

（一）基于机器学习的运河文化资源分类与标注算法

在图像处理领域，深度学习技术中的图像分类模型能够自动识别和标注建筑类型、装饰纹样、建筑材料等关键信息，实现对运河沿线建筑文化遗产的精细化分类管理。卷积神经网络的应用使得原本需要专家耗费数月完成的标注工作能够在短时间内高效完成；在文本处理方面，文本分类算法能够对档案、契约、文学作品等文献资源进行自动分类，同时标注主题、情感、写作风格等深层语义信息；多模态学习正逐渐成为研究热点，通过整合图像、文字、3D模型等多种模态数据，实现更精准的分类效果和更丰富的自动化标注内容。

（二）运河文化时空演变模式的大数据分析方法

通过数字化技术重建不同时期的地图资料、文本记录、考古发现，能够系统复原运河及其沿岸聚落的历史变迁轨迹。将GIS技术与时态数据相结合，从四维视角审视文化现象的演变过程。构建以城市为节点、运河为边的交通网络模型，运用网络分析方法，通过计算中心度、聚集系数等指标量化不同时期城市地位的变化趋势。借鉴流行病学的传播模型，研究建筑风格、工艺技术等文化要素沿运河的传播路径和扩散机制。

（三）跨文化运河遗产关联性挖掘与知识发现机制

使用关联规则挖掘方法探索文化元之间的关联，例如文化元之间“正外部性”的，即建筑形式、风格与GDP的相关性关系；一些节日民俗与宗教的相关性关系；等等。其中Apriori、FP-Growth等是关联规则分析经典算法在海量的文化数据挖掘中能有效解决分析问题的典型案例。聚类分析算法对一批具有相似的文化属性的对象进行聚类自动归纳，可挖掘和发现被低估的文化相关性和传播路径^[3]。异常检测能够发现样本数据中与平常数据显著不符的异常，并且可以推测该异常样本的文化创新和外来文化融入的可能；根据已有的历史资料数据和演化规律去预测历史资料中的缺失部分。

（四）运河文化资源价值评估指标体系与量化分析模型

建立历史性、艺术性、科技性、社会性等全方位的评价指标体系。历史性评价指标有文物的年代、历史文化独特性等；艺术性评价指标有其审美的风格、艺术价值等；科技性评价指标有科

研价值及科技创新性等；社会性评价指标有当今社会功用的发挥及社会认知度。

（五）基于大数据的运河文化传承短板精准研判分析

传统运河文化传承与保护工作多依靠经验判断，存在保护重点模糊、短板识别不准、资源利用不均衡等问题，难以实现精准化、精细化保护。借助大数据数据爬取、数据统计与态势分析技术，可全面采集运河沿线文物保护现状、文旅开发数据、非遗传承人数、网络传播热度、民众关注度、文献引用频次、景区客流量等多维度数据。通过数据横向对比、纵向分析，精准研判不同河段、不同城市的运河文化资源保护现状、传承现状与开发短板。例如可通过热度数据对比，发现部分高价值小众非遗、古遗存曝光度低、传播力度不足的传承短板；通过保护数据比对，识别部分河段重开发、轻保护，文物修缮不及时、资源闲置浪费等现实问题；通过用户行为数据分析，掌握当代大众对运河文化的认知偏好与接收短板。依托大数据研判结果，能够精准区分优势资源、弱势资源、濒危资源，针对性定位传承薄弱环节，为制定差异化、精准化的运河文化保护传承策略提供数据支撑，有效避免传统保护工作的盲目性与同质化问题。

（六）大数据赋能运河文旅融合创新与资源活化应用

在文旅融合发展背景下，运河文化资源的活化利用是文化传承的重要途径，大数据技术可为运河产业化、场景化、生活化应用提供全新支撑。通过汇总分析全网文旅消费数据、游客出行数据、文旅偏好数据、短视频传播数据、景区消费画像数据，可精准把握当代文旅市场消费趋势与大众文化需求。结合运河沿线各类文化资源的价值量化结果，能够精准筛选具备文旅开发潜力的特色文化元素，如运河漕运文化、水工文化、民俗文化、古镇文化、诗词文化等。同时依托用户画像分析技术，精准匹配不同年龄、不同群体的文旅需求，针对性打造沉浸式文旅场景、特色文创产品、文化研学线路、数字展览内容，打破传统运河文旅产品单一、内容同质化、体验感不足的弊端。此外，通过大数据实时监测文旅运营数据、游客反馈数据、文化传播数据，可动态优化运河文旅开发模式与传播路径，持续提升运河文化的传播力、影响力与市场活力，实现运河文化资源从静态保护、档案留存，向动态传承、活化利用的高质量转化。

三、智能化运河文化资源数据库服务平台构建与应用

（一）面向多终端的运河文化资源智能检索与推荐系统

准确理解用户的深层信息需求是构建智能检索系统的核心要素。传统的关键词匹配方式难以充分满足用户的查询意图，基于语义理解的智能检索技术能够识别同义词、上下位词关系，并推断隐含的查询意图^[4]。例如，当用户检索“运河古镇”时，系统除了返回直接匹配的结果外，还能够智能推荐“历史街区”“水乡民居”等相关内容。个性化推荐算法通过分析用户的浏览历史、收藏记录、页面停留时间等行为数据构建用户画像，运用协同过滤技术发现相似用户的兴趣偏好，内容过滤技术分析资源自身的特征属性，混合推荐策略整合多种算法的优势以提供更精准的推

荐结果。跨语言检索功能突破了语言障碍,考虑到运河文化的国际性特征和资源描述的多语种特点,机器翻译与跨语言信息检索技术使得用户能够使用母语检索并获取多种语言的文化资源,这不仅便于研究者开展国际比较研究,也使普通公众能够接触到更丰富的文化内容。移动端界面设计充分考虑不同使用场景的需求,游客在运河景点可能更需要地理位置相关的导览信息,而研究者则更关注深度的文献资料和学术资源。

(二) 基于 GIS 的运河文化遗产可视化展示与交互设计

虚拟复原往昔运河实景,结合激光测距扫描、倾斜摄影测量获取的点云信息及以往的相关历史地图、考古文献等多源数据,构建往昔不同时代的运河景观,可让用户进行虚拟漫游,亲身体验运河的历史发展演变过程,相较于单纯图示文字解说更具感染力。采用时空地图的记录手法叙述运河发展历史,将不同时代背景下的事件、人文活动、文化现象与地图界面相互叠加,通过控制时间轴演示不同历史时期的样貌变化,跟随古时商人的轨迹线路去了解运河上各处贸易分布情况,可见某一创新发明产品是怎么顺着运河路线进行扩散的,不同时间段下的文化变迁显而易见。利用热力图展示文化资源的空间聚集程度,哪里有更多数量的遗产,哪里某一类型遗产空间分布较为广泛,通过颜色深浅的视觉表达更加快速有效地完成信息的传递。可视化的表达方式更利于受众高效全面掌握运河文化整体的布局情况,也为文化资源的保护规划提供合理决策支持。运用 AR(增强现实技术)将虚拟融入现实,游客手持手机对眼前遗址进行扫描即可将该遗址往昔真实实景模型显示在画面当中^[5]。

(三) 运河文化资源开放共享机制与数据安全保障体系

实现开放共享的技术基础是开放应用编程接口 API(open application programming interface)的标准化设计,任何用户单位如科教文企事业单位可以通过统一 API 接口调用开放数据并提供具有针对性的专业应用服务。统一开放共享 RESTful API 的设计应具有简洁易用的属性,分组不同的访问权限即可以满足任何类型用户的需求。数据提供单位应当制定统一的共享协议以平衡对外开放共享与知识产权的关系,统一使用国际上常见的 Creative Commons 等许可证协议以规定数据资料开放共享的类型与授权使用。基础性数据资料免费开放获取,对核心研究结果应当注明出处来源,对商业性的使用需要提出相应授权。数据分层

次管理,可以达成开放共享知识的同时保护成果知识的权利,做到张弛有度的平衡关系的处理。区块链技术可以较好地解决数据的确权问题,对所有数据的提供方信息、变更数据信息等均能形成永久记载在区块链数据账本且不能篡改的数据档案,这样就可以达到数据可追溯可核实的效果。例如名贵文物的数字化资料,数据资料可以通过区块链数据证书提供权威的认证资料。从网络安全到应用数据安全再到数据自身安全,可以分层级进行综合性的数据安全防护。系统要设防火墙、入侵检测系统等保障网络安全,通过严格的身份验证机制、权限授权机制保障应用层数据安全,对数据自身还可以从加密手段、定期备份机制等进行数据安全保障。定期进行系统安全审计和应急处置演练以提升系统的安全管控能力。

(四) 数据驱动的运河文化传承创新与决策支持应用

数字运河文物所具有的数据属性,使得历史数据和环境监测数据的输入、输出是可能的;进行模拟或仿真分析,以预知特定文物受到的破坏情况,如古遗址地下水位变动对其地基的影响、遗址核心区或老城区改造对历史环境所造成的冲击、气象灾害对户外文物本体的伤害等,通过仿真计算,实施不同的文物保护方案,作为进一步确定文物保护方案的依据;在文物保护方面,已有的虚拟修复技术可以利用计算机系统来模拟文物修复后的新旧效果比较,在文物修复前进行修复方案的模拟,与建筑仿真 CAD 软件的设计类似^[6]。

四、结语

大数据技术为世界运河文化资源的保护传承开辟了新的发展路径,从分散到整合、从静态到动态、从保存到活化,数字化转型正在深刻地重塑着文化遗产工作的格局与模式。然而必须认识到,技术始终是实现目标的手段,文化本身才是核心和目的。在推进数字化建设的过程中,必须始终把握文化遗产的本质属性和核心价值。数据库不应该成为缺乏温度的数字存储空间,而应该构建成富有生命力的文化传播平台,让运河的故事被更多人知晓,让运河的智慧在当代焕发新生。这正是技术赋能文化的根本意义和价值所在。

参考文献

- [1] 张涛,周卫.广西非遗文化知识图谱构建与数据处理研究[J].智能计算机与应用,2025,15(3):72-78.
- [2] 李鸿雁.中国土司世界文化遗产文献资源网络数据库的构建[J].吉首大学学报(自然科学版),2024,45(4):92-96.
- [3] 张希,蒋鑫,张诗阳,王向荣,林箐.大运河文化遗产利用的公众感知研究——基于网络数据的语义分析[J].中国园林,2022,38(1):52-57.
- [4] 蒋鑫,张希,钱行健,林箐,王向荣.基于多源数据的运河历史文化街区(名镇)“真实性”感知评价与更新对策研究[J].现代城市研究,2021,36(7):20-27+37.
- [5] 徐芳,李亚宁.大运河线性文化遗产数字资源的关联数据模型构建及其实现[J].图书馆学研究,2022,No.521(06):45-53.
- [6] 王燕燕,王浩,唐晓岚.中国大运河遗产构成及遗产价值研究综述——基于 CNKI 中国知网远程数据库[J].南京林业大学学报(人文社会科学版),2019,19(03):70-83.