

Exploration of the Application of Big Data Technology in the Development and Utilization of Archival Information Resources

Zhiqiang Zhu

Handan Transportation Construction Investment Management Center, Handan, Hebei, 056000, China

Abstract

The rapid development of big data technology has opened up a new path for the development and utilization of archival information resources. This article starts from the background of technological applications, systematically reviews the practical progress of big data technology in archive data integration, value mining, and service innovation, and analyzes the current bottlenecks faced. Research has found that although big data technology has significantly improved the digital conversion efficiency and intelligent analysis capabilities of archival resources, its application is still limited by issues such as data security risks, insufficient technological adaptability, and outdated management mechanisms. Based on this, this article proposes strategies such as building a multidimensional technology collaboration system, strengthening data governance framework, and improving talent pool construction to promote the sustainable development of archive information resource development and utilization. The research aims to provide theoretical support and practical reference for technological innovation in the field of archives.

Keywords

big data technology; Archive information resources; Data governance; Technical collaboration; Intelligent services

大数据技术在档案信息资源开发与利用中的应用探索

朱志强

邯郸市交通建设投资管理中心, 中国·河北 邯郸 056000

摘要

大数据技术的迅猛发展为档案信息资源开发与利用开辟了全新路径。本文从技术应用背景切入,系统梳理了大数据技术在档案数据整合、价值挖掘及服务创新中的实践进展,并剖析了当前面临的现实瓶颈。研究发现,尽管大数据技术显著提升了档案资源的数字转化效率与智能化分析能力,但其应用仍受限于数据安全风险、技术适配性不足及管理机制滞后等问题。基于此,本文提出构建多维度技术协同体系、强化数据治理框架、完善人才梯队建设等策略,以推动档案信息资源开发与利用的可持续发展。研究旨在为档案领域的技术革新提供理论支撑与实践参考。

关键词

大数据技术; 档案信息资源; 数据治理; 技术协同; 智能化服务

1 引言

数字化时代背景下,档案信息资源作为社会记忆与知识资产的核心载体,其开发与利用效能直接影响文化传承与公共服务的质量。传统档案管理模式因技术局限,难以应对海量数据的存储、检索与分析需求,导致资源利用率低、服务模式单一。大数据技术通过数据挖掘、机器学习等技术手段,为档案资源的深度开发提供了技术支撑,如文本语义分析助力历史档案的语义重构,用户画像优化档案服务的精准性。然而,技术应用过程中存在数据安全隐患、跨平台兼容

性不足、人才储备短缺等矛盾,亟需系统性解决方案。

2 大数据技术赋能档案信息资源开发与利用的动因与意义

2.1 技术驱动下的档案管理范式转型

数字社会的演进进程加速发展,档案信息资源库的规模正以指数级形式迅速扩张,现行的人工整理与分类方法已无法跟上高效管理的步伐,大数据技术基于自动化数据采集、分布式存储与智能分析进行数据挖掘,实现了档案资料从生成到终结的全面生命周期控制。某省档案馆采纳了Hadoop架构体系结构,档案检索效率增长八成以上,资源运用效率显著攀升,现代信息技术的应用极大地提升了数据存储的经济效益,明显促进了档案服务从“被动应询”向“主

【作者简介】朱志强(1979-),男,中国河北魏县人,本科,助理馆员,从事档案管理研究。

动推送”模式的过渡，提高了档案处理流程的效能^[1]。

2.2 档案资源社会价值的深度释放

档案信息资源富含深厚的历史、文化和经济价值内涵，其长期处于“休眠”状态，严重限制了其社会效益的充分发挥，大数据技术借助多模态数据分析手段，实现了对文本、图像、音视频等数据的跨模态关联，拓展了数据处理的边界，深入挖掘档案资料的潜在联系及其潜在的应用领域。

3 大数据技术在档案开发与利用中的现实挑战

3.1 数据安全隐患与技术伦理风险

档案资料中积聚了大量涉及个人隐私及政府机密的敏感信息，开放资源的共享应用需平衡安全防护与信息共享的关系，在数据资源挖掘的大数据环境里，数据流动的跨平台特性日益明显，使得数据泄露的风险指数级上升，某机构因数据脱敏处理不周全，引发用户信息泄露事件，算法偏见可能引起对历史档案的误读与扭曲，有必要构建一套技术伦理的审查体系。

3.2 技术适配性与资源整合困境

异构数据整合所遭遇的技术难题分析，档案资料的来源形式多样，包括纸质档案、电子文件以及社交媒体等多种信息记录媒介，格式不一致性引发了数据清洗及整合过程中的成本大幅攀升，某高等学府的数字化项目面临元数据标准缺失的困扰，导致30%的档案资料无法实现有效对接。

技术手段与业务需求之间存在显著的不匹配现象，在技术工具的引入方面，部分档案机构存在盲目跟风现象，忽略了业务场景与实际应用场景的匹配要求，某档案馆所采用的分类系统因数据训练资源匮乏而受限，评估误差比率之高，高达40%。

3.3 管理机制与人才储备滞后

档案管理机构内部组织架构与信息技术应用之间存在显著的不匹配现象，暴露出部门间协同机制的缺陷，档案学领域，既通晓档案学理论又精通技术操作的复合型人才培养有限，限制技术实施效果之发挥，某地档案局经过周密分析得出，基础数据分析技能的掌握率在员工中仅为15%。

4 大数据技术赋能档案开发的优化策略

4.1 构建多维度技术协同体系

档案资源开发需深化大数据技术的应用，突破其单一技术的边界，运用多技术维度协同作用，达成数据深度整合、风险实时监控及跨平台互动协作。

4.1.1 异构数据融合技术

档案信息资源呈现出来源多样、格式结构各异的显著特性，现有技术手段在处理文本、图像、音视频等多模态数据整合方面存在明显的技术瓶颈，实施语义标注及本体建模技术手段，可对非结构化档案资料实施语义分析与关联性重建，构建一个综合性的知识表述框架。国家图书馆运用资源描述框架(RDF)技术对古籍文献及数字档案实施语义标记，

研发跨库数据关联性模型实施路径，开发“关键词—实体—事件”的多层级语义检索技术，用户检索效率增长六十个百分点，进一步强化知识图谱技术支撑，将零散的档案资料整合构建为系统化的知识网络体系，主张动态推理技术的应用与场景化实施的理论拓展。以某省级档案馆为分析对象，采用历史档案资料构建知识图谱实施步骤，探讨清末商埠贸易与当代区域经济地理关联的互动关系，为政策制定提供历史背景资料支撑背景分析，该技术的核心机理在于运用语义映射技术来化解数据异构性问题，应同步研制针对档案特性而设的专用标注装置，采取手段以降低人工标注作业的经济开销^[2]。

4.1.2 动态风险评估与防护

档案资料共享与安全保障的平衡策略研究，区块链技术的分布式账本机制确保了档案数据操作的全程可追溯与不可篡改性，某国际档案协作组织采用Hyperledger Fabric技术架构搭建了联盟区块链，对档案查阅、编辑等操作实施实时上链并完成存证，降低数据篡改风险至90%，提高档案安全防护水平。基于联邦学习的技术手段，可以在不共享原始数据的基础上，完成多方的联合模型训练，跨部门合作中敏感档案资料的综合分析创新模式探索，某市档案管理机构与高等教育机构的研究力量携手合作，采用联邦学习技术对民国时期户籍档案进行数据统计与阐释，在确保公民隐私安全的前提下，深入开展人口迁移规律的研究工作。应当研制一套动态化的风险评估模型架构，运用先进技术对数据传输过程中的异常情况进行实时监控，某国家级档案信息平台部署了一套先进的基于机器学习的人侵检测体系，对用户行为模式进行深入挖掘，高效防御系统成功遏制了98%的非法入侵尝试，技术融合的核心要素是构筑“区块链技术、联邦学习与人工智能监测”相结合的综合安全防护架构，全面涵盖数据存储、传输及应用的各个应用场景。

4.2 强化数据治理框架

档案资源的持续开发离不开数据治理的支撑，必须通过实施标准化工程和全方位流程监管来提升数据质量及合规性水平。

4.2.1 标准化与合规性建设

档案信息的多样性及其历史属性引发了元数据规范的缺失与互操作性的欠缺，欧盟颁布的《文化遗产数据共享指南》采纳了统一的都柏林核心元数据(DC)与欧洲数据模型(EDM)标准，实施跨国界档案资源的语义对接计划，此经验对中国具有借鉴意义。中国应迅速启动档案大数据国家标准的相关制定程序，确立数据搜集、阐述及交流各阶段的技术标准规范，实施民国档案资料数字化行动，南京某档案馆携手高等院校共同研发《近代文献元数据标准化规范》，整合文件分类、时间编码机制、人物关系模式等十二个核心信息范畴，将跨库检索的精确度由72%提高至89%。有必要构建一套数据资源共享的等级分类体系，对公开、受限与

机密数据实施分层次访问权限管理,该省级政务档案系统采纳了“三色标识”的管控模式,包括绿色标识代表公开信息、黄色标识表示需申请授权信息以及红色标识表示禁止访问信息,在维护数据安全的前提下,将数据开放的比例提升至45%。

4.2.2 全流程质量控制

自数据采集至服务输出的整个操作序列,档案开发效能的优劣受质量控制环节的直接影响,在数据采集阶段,运用光学字符识别技术中的纠错算法,可以有效提高纸质档案数字化处理的精确度,清华大学自主开发的古籍智能校对系统,融合了卷积神经网络技术及规则库,明清档案鉴定技术升级,错误率降至3%。在存储环节,必须实施冗余校验及版本管理技术,某市级档案馆启动了星际文件系统(IPFS)分布式存储网络的实施进程,运用哈希算法对数据进行校验,以保障数据的完整性,需纳入人工智能驱动的品质检验模块,某历史档案馆借助自然语言处理技术,对档案转录文本中的时序冲突与逻辑谬误进行自动识别,数据可靠性水平较之以往提升了40%,全流程质量控制的核心挑战在于构建适用于各类档案的校验规则体系,需融合专业学者见解与机器学习技术,对检测模型实施动态优化调整^[9]。

4.3 完善人才梯队与协同机制

技术实施之成效,依赖于人才与组织协同作用的实现,必须打造“教育、实践、创新”三位一体的综合支撑架构。

4.3.1 跨学科人才培养

档案管理人员普遍面临技术能力上的局限,我们高等教育课程体系改革是推动教育改革向纵深发展的必然选择,推出“档案大数据分析”与“数字人文技术”等综合性交叉课程,培育具备档案学理论素养及Python、SQL等信息技术应用技能的复合型人才。上海市档案局携手复旦大学共同打造“智慧档案管理”专业教育品牌,课程内容丰富,包括数据清洗、可视化分析以及区块链存证等多个领域,高校毕业生就业领域匹配度达到85%,就业市场前景看好,企业实训项目是知识转化加速的推动力,某科技企业与档案馆携手实施“人工智能档案分类”实战演练项目,学员借助实际数据集进行模型优化操作,学业完成后,独立开发了一套档案自动化索引系统。需形成一套持久的终身学习体系,某省份档案馆实施“技术学分制”管理模式,图书馆工作人员需

每年完成40学时的大数据专题培训课程,将考核成绩与职称晋升路径相绑定。

4.3.2 产学研协同创新

档案部门与技术服务商之间的供需不匹配是导致工具应用效率低下的根本原因,搭建协同创新实验室,便于需求侧与供给侧开展直接交流,中国首屈一指的历史档案馆与阿里云公司共建“档案智能计算实验室”,针对明清档案满文本,研发特定的光学字符识别平台,识别精度由68%显著提高至92%。专项研发策略能够有效消除业务中的核心痛点,某企业为档案馆量身打造的“智能编目软件”成功整合了本体推理技术,自动生成档案主题词汇并赋予相应的分类编号,每日编目数量由200件增至600件,编目效率显著提高,应当打造跨组织合作平台,长三角地区档案大数据联合体依托共享的算力资源和算法模型进行协作,削减中小型档案馆信息技术应用的经济负担,实现协同创新的核心任务是构建知识产权共享与利益分配的框架,催生多元主体参与积极性。

5 结语

大数据技术为档案信息资源的开发与利用注入了变革性动力,但其应用需突破技术、管理与伦理的多重壁垒。本文从技术协同、数据治理与人才培育三方面提出系统性策略,强调通过标准化建设、动态风险防控及跨领域协作,实现档案资源的安全共享与价值增值。案例表明,技术工具与业务场景的深度融合是提升档案服务效能的关键。未来,随着边缘计算、知识图谱等技术的深化应用,档案信息资源有望从静态“仓库”转型为动态“知识引擎”,为文化遗产与社会创新提供持续动能。研究进一步呼吁行业关注技术伦理与用户需求,推动档案开发从“技术主导”向“人本导向”演进。

参考文献

- [1] 徐文静.运用大数据技术加强干部人事档案资源体系建设探析[J].信息与电脑(理论版),2024,36(14):189-191.
- [2] 琚小琳.基于大数据技术的固定资产投资项目档案信息分析与价值挖掘[J].市场瞭望,2024,(13):41-43.
- [3] 段竹莹,牛相林.基于大数据技术的电子档案管理现状和发展对策[J].中关村,2023,(04):106-107.