Modern Management 现代管理, 2024, 14(11), 2837-2841
Published Online November 2024 in Hans. https://www.hanspub.org/journal/mm
https://doi.org/10.12677/mm.2024.1411337

Hans汉斯

人工智能视域下的档案资源开发利用研究

黄莹婕

扬州大学社会发展学院，江苏 扬州

收稿日期：2024年10月8日；录用日期：2024年10月18日；发布日期：2024年11月12日

摘　要

随着互联网的广泛普及，档案资源的使用者对档案开发的需求正日益增长。作为全球科技创新的焦点，人工智能技术正被众多学科领域密切关注，并被有效融入各自的研究与实践之中，取得了诸多成果。因此，档案资源的开发工作也应紧跟时代步伐，充分利用人工智能技术来推动档案事业高质量发展。笔者通过归纳总结人工智能赋能档案资源开发的典型技术，进一步分析人工智能应用于档案资源开发中面临的困境，最终针对性地提出了优化策略。

关键词

人工智能，档案资源开发，智能化

Research on the Development and Utilization of Archival Resources in the Context of Artificial Intelligence

Yingjie Huang

School of Social Development, Yangzhou University, Yangzhou Jiangsu

Received: Oct. 8th, 2024; accepted: Oct. 18th, 2024; published: Nov. 12th, 2024

Abstract

With the widespread popularity of the Internet, the demand of the users of archive resources for archive development is growing. As the focus of global technological innovation, artificial intelligence technology is closely monitored by numerous disciplines and effectively integrated into their research and practice, achieving many results. Therefore, the development of archival resources should also keep pace with the times and fully utilize artificial intelligence technology to promote high-quality development of the archival industry. The author summarizes typical technologies for

文章引用: 黄莹婕. 人工智能视域下的档案资源开发利用研究[J]. 现代管理, 2024, 14(11): 2837-2841.
DOI: 10.12677/mm.2024.1411337

empowering archive resource development with artificial intelligence, further analyzes the challenges faced by the application of artificial intelligence in archive resource development, and ultimately proposes targeted optimization strategies.

Keywords

Artificial Intelligence, Development of Archival Resources, Intelligentization

Copyright © 2024 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

人工智能，简称 AI，其概念由麦卡锡首次提出。人工智能致力于探索、构建并优化一系列理论、方法、技术以及应用体系，旨在模拟并拓展人类智能的边界，是一门综合性的技术科学[1]。近年来，人工智能技术取得了显著进展，其应用范围持续扩大，逐渐成为推动科技变革的关键因素。2017 年，我国正式将人工智能提升至国家战略层面。紧随其后，在 2018 年召开的中国档案学会年会上，关于人工智能的主题演讲受到了档案学界的广泛瞩目。至 2021 年，《“十四五”全国档案事业发展规划》明确指出，应积极探索知识管理、人工智能、数字人文技术在档案信息深度加工与利用中的应用，并加强大数据、人工智能等新一代信息技术在数字档案馆(室)建设中的融合应用，旨在通过人工智能促进档案管理工作的转型升级，与数字档案馆(室)建设紧密结合[2]。2022 年度，国家档案局在科技项目的立项工作中，着重突出了对人工智能技术融入档案管理领域的深入探究需求。2023 年，国家档案局科技项目立项选题依旧将“人工智能在档案管理中应用的现状与前景研究”作为重点项目。由此可见，人工智能技术在档案资源管理与开发中的应用已受到学界与实务部门的普遍重视。对人工智能应用条件的深入研究，将有助于充分发挥其优势，提升档案管理与开发的效能。

2. 人工智能技术赋能档案资源开发利用

2.1. 智能语音技术

档案资源涵盖了静态的纸质档案、电子档案以及动态的音视频档案等多种类型。相较于文本档案，音视频档案在利用过程中面临着诸多困境，智能语音识别技术的应用为解决这些问题提供了有效路径。智能语音技术旨在实现人与机器之间以语言为媒介的交流，伴随着深度神经网络的应用，语音识别的准确性得到了显著提升，并于 2016 年首次超越了人类水平，这标志着智能语音识别技术进入了实用阶段。智能语音识别系统的运作流程包括五个关键环节：语音信号的采集、预处理、特征参数提取、核心识别以及后期处理[3]。通过这些流程，音视频档案能够实现自动翻译并匹配字幕，或进行文本转录，从而极大地拓宽了其利用范围，进一步丰富了档案资源。

2.2. 光学字符识别技术

光学字符识别(OCR)技术是一种通过计算机扫描识别文字，并将其转化为数字化形式的技术。其核心运行流程包括影像采集、预处理、特征提取、比对识别、人工校正及结果输出六大步骤[4]。OCR 技术具备将海量纸质档案迅速转化为电子文本的能力，显著缩减了档案数字化的周期。相较于传统的手工录入方式，它能够大幅提升工作效率并减少人力成本支出[5]。OCR 软件的识别精度较高，确保了检索结果的

准确，进而提升了档案的实用价值。目前，在遇到书写规范、图像清晰的文本时，OCR 技术能够实现高准确率的转录。然而，在面对复杂的古籍文献档案时，转录的准确性往往会因遇到古旧繁复的文字而受到影响。

2.3. 语义分析技术

档案智能检索系统采纳了人工智能领域的语义工程技术，借助机器的自然语言理解能力来优化档案检索过程。当用户输入自然语言作为检索关键词时，该系统并非仅仅进行机械的词汇匹配来输出结果，而是能够智能地关联并展示更多与检索词相关的词汇，以满足用户可能存在的多样化需求。这一基于语义的档案智能检索系统涵盖了五个关键环节：首先，从各类数字化档案中提取字符串；其次，基于概念对文本进行分词处理；接着，构建基于本体的档案语义信息体系；然后，对档案内容进行语义标注；最后，实现语义查询。通过上述流程，系统能够对用户的检索词进行深入的语义分析，构建概念与数字档案之间的关联，从而推动档案资源实现更深层次的开发与利用。

3. 人工智能应用于档案资源开发的困境

3.1. 档案资源种类繁多导致机器识别面临困难

档案通过 OCR 技术实现数字化，极大地提升了档案信息的可利用性。然而，档案资源的多样性、复杂性以及记录符号的差异性，如方言档案或少数民族文字档案，极大地增加了文本检测与识别的难度。此外，档案中图像、表格、图表、图形等记录方式的存在，也对机器的识别过程构成了障碍[6]。在档案著录环节，由于早期档案格式的不规范，信息位置的不统一，OCR 识别在提取有用信息时面临困难，导致机器学习在覆盖所有记录类型时易发生误差，仍需耗费大量时间与人力进行校正。

3.2. 档案智能平台在用户隐私保护方面意识薄弱

随着网络技术的迅猛进步，用户信息的搜集与运用变得愈发高效。在大数据浪潮的席卷下，人工智能凭借对多元数据的深度分析，能够洞察用户的浏览习性，据此生成预测性信息，为用户提供个性化的服务体验。然而，这项技术如同双刃剑，既带来了前所未有的便利，也潜藏着用户信息泄露的风险。目前，档案信息资源服务平台的建设普遍采用外包合作方式，这导致用户在平台活动期间产生的诸如用户名、联系电话、IP 地址、身份证号等敏感信息，有可能被外包服务商所采集。在人工智能渗透至档案资源开发领域的进程中，对用户隐私安全的保护意识尚显薄弱，已浮现的档案信息安全隐患，对档案信息化建设的步伐构成了显著阻碍。鉴于此，为确保人工智能技术在档案资源开发领域的广泛运用得以安全无虞地推进，必须配套实施先进的保密技术手段，以切实维护用户隐私权益，从而推动档案服务向更高品质发展。

3.3. 档案服务在人工智能技术运用上尚显不足

在我国，人工智能正致力于推动档案的智慧化发展，旨在简化档案工作流程，提高效率。然而，在利用方面，人工智能的应用仍较为有限，与用户之间的联系不够紧密。部分省级档案馆网站的数字档案资源缺乏深入开发与加工，用户可获取的资源有限，且大多侧重于政务档案，民生档案与特色档案相对较少[7]。此外，我国省级档案馆的服务内容参差不齐，较多档案网站在人工智能技术的运用上较为匮乏，如知识库技术、专家系统、模糊信息处理技术、数据挖掘技术以及智能平台等，导致档案资源无法有效地向用户展示，影响了档案服务利用的效果。因此，面对用户利用档案需求的多样性，人工智能在档案服务利用上的开发与优化仍需进一步加强。

4. 人工智能应用于档案资源开发的优化策略

4.1. 关键技术聚焦，加快核心技术研究

目前，科大讯飞等企业在语音识别领域取得了显著进展，已能将口述内容同步转化为文字，并在民国时期繁体文书档案的识别中取得了实际应用效果[8]。基于深度学习的 OCR 技术可通过卷积神经网络、循环神经网络 RNN 等手段解决传统 OCR 技术中的形似字判断难题。在著录方面，工作人员可直接从 OCR 中提取相关信息，并复制粘贴至目录数据库的相应位置。同时，加快核心技术研究算法优化与创新。对现有算法进行优化和改进，提高处理效率和准确率。探索新的算法模型，如迁移学习、联邦学习等，以适应档案馆数据的多样性和分布性。研究如何高效处理和存储海量档案数据，包括分布式存储系统、并行计算框架等。并且开发适用于档案馆的大数据分析工具，实现档案信息的深度挖掘和价值发现。

4.2. 用户信息加密，降低隐私泄露风险

为了深入理解档案用户的需求并提供服务，应用人工智能必须广泛收集与分析用户行为数据。尽管这一过程极大地促进了档案服务的便捷化与精准化，但同时也对用户隐私安全构成了潜在威胁，引发了隐私保护方面的风险[9]。为了应对此风险，首先，应严格遵守相关法律法规，如新《档案法》中关于知识产权和个人隐私保护的规定，以及《大数据保护法》《个人信息保护法》《著作权法》等相关法律，在档案开放利用过程中均应严格执行。其次，档案智能平台在提供服务时，应对用户 IP 地址、电话、身份证号等敏感信息进行自动加密处理。电子档案经过加密后，符合条件的用户可直接访问其内容，从而确保档案信息及用户信息的传输安全。最后，对于涉密档案，应采用加密技术，并结合云计算与机器学习技术，识别并分类电子档案中的敏感信息，根据不同密级设定差异化的开放利用时间与范围[8]。

4.3. 强化政府政策引领，建立健全监管体系

《国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要》强调了科学创新体系的完善需求，而新修订的《档案法》亦倡导在档案利用中采纳创新技术。鉴此，档案部门在推进人工智能融入档案开发与利用的过程中，亟需制度层面的有力支撑。国家层面的档案管理机构可采取一系列举措，以构建人工智能创新体系，包括但不限于：定期发布人工智能创新年度规划、增加科研经费的专项投入、强化人工智能创新平台的搭建、吸纳顶尖科技人才、设立专项技术项目、深入实施前瞻性的技术探索与应用实践。这一系列行动应自上而下地贯穿于各级档案馆的工作之中，旨在实现档案开发与利用工作和人工智能技术的深度融合。此外，还需构建科学合理的激励与考核机制，合理配置人员岗位，以确保创新体系与机制的有效实施与持续优化。

5. 结语

人工智能技术的兴起为档案资源的开发利用开辟了新机遇，同时也带来了诸多挑战。人工智能的应用加速了档案的数字化进程，并推动了档案平台的智能化升级。然而，现实层面的制约因素限制了人工智能在档案领域的广泛应用。因此，可以通过聚焦关键技术、加密用户信息、强化政府政策领导等措施，优化其在档案开发利用的应用，以期更多档案实现价值最大化。

参考文献

[1] 王秋洁, 孙军. 综合性档案馆应用人工智能创新管理的 SWOT 分析[J]. 山西档案, 2020(3): 148-152.

[2] 本刊讯. 中办国办印发《“十四五”全国档案事业发展规划》[J]. 中国档案, 2021(6): 18-23.

[3] 李雪林. 基于人机互动的语音识别技术综述[J]. 电子世界, 2018(21): 105.

[4] 迟春佳. OCR 技术及其在高校图书馆信息资源数字化建设中的应用[J]. 中国科技信息, 2007(7): 95-96.
[5] 王瑜. 对 OCR 识别技术在档案管理应用的一点思考[J]. 卷宗, 2021, 11(4): 173.
[6] 李欢, 施瑞婷, 张洁. 人工智能在档案资源开发利用中的应用分析[J]. 山西档案, 2021(2): 71-78.
[7] 刘青, 王雯洁. 省级档案网站档案知识服务研究[J]. 档案天地, 2020(11): 28-30, 59.
[8] 郑慧, 刘思含. 人工智能与档案开发利用: 应用、愿景与进路[J]. 山西档案, 2022(5): 5-10+28.
[9] 岳幸晖, 杨智勇. 人工智能在档案管理中的应用图景与风险防范[J]. 档案与建设, 2023(10): 36-40.