

Present Situation and Forecast of Intelligent Information Technology Application in Archives

Xianghong Cui ^{1,a,*}, Bin Gao ^{2,b}

¹ Shandong Technology and Business University, Yantai, China

² Shandong University of Technology, Zibo, China

^a cxh_sd@163.com, ^b gaobin_sd@163.com

* Xianghong Cui

Keywords: Smart Archives, Intelligent Archives, Artificial Intelligent, Artificial Intelligent Technology.

Abstract. With the rapidly developing of artificial intelligent technology (AIT) in CHINA, the archives work is stepping into a new stage of intelligent archives. The Second Information and Intelligent Archives Forum and Academic Seminar held successfully in 2018, which brings new opportunities and challenges to the research on intelligent archives. This paper first presents the scope of AIT, then analyzing the situation of AIT's application in archives work, Finally, some suggestions on archive's development has been proposed.

智能信息技术在档案中的应用现状及展望

崔向红^{1, a,*}, 高斌^{2, b}

¹山东工商学院, 烟台, 中国

²山东理工大学, 淄博, 中国

^a cxh_sd@163.com, ^b gaobin_sd@163.com

*崔向红

关键词: 智慧档案、智能档案、人工智能、人工智能技术

中文摘要. 人工智能技术在中国快速发展, 引领档案工作进入智慧发展阶段。2018年第二届信息化与智慧档案论坛暨学术研讨会成功召开, 对档案界研究智慧档案带来了新的机遇和挑战。本文首先对人工智能技术的范畴加以明确, 然后对人工智能技术在档案领域的应用现状做了分析, 最后针对国内人工智能技术在档案工作中的应用做出前瞻、提出建议。

1. 引言

信息化时代, 大数据、云计算、物联网、移动互联和人工智能技术快速发展。《国家信息化发展战略纲要》和《全国档案事业发展“十三五”规划纲要》对信息化工作均提出了明确要求。智慧档案(馆)以新一代信息技术的运用为基本手段, 以提高档案管理和基本目标, 是“智慧”背景下档案馆发展进化的高级形态。在当前信息化技术不断发展和应用的前提下, 智慧技术和智慧管理已成为新的发展趋势, 并为档案信息化的未来发展提供了新理念、新思路和新目标。

2. 智能信息技术在档案领域的范畴与内涵

人工智能技术，也称智能信息技术是实现智慧档案应用的技术基础。因此，有必要理清其范畴与在智慧档案中的内涵。

2.1 智能信息技术范畴

近年来人工智能技术的发展使其在各领域的智慧应用研究空前热烈，但所说的人工智能技术到底指哪些？具体说来，却又不甚了了。有人说，大数据是人工智能；有人说，云计算是人工智能；还有人说物联网是人工智能。

这些说法其实都不准确，也就是说既对又不对。谭铁牛院士在全国院士大会的报告中说“人工智能领域的误解和炒作普遍存在”、“存在有意地炒作并通过包装人工智能概念来谋取不当利益”的现象[1]。也正是这些有意无意的误导，混淆了人们的视听，混乱了人们对人工智能的理解，使得人工智能的概念更加模糊起来。因此，通过对当前人工智能技术追根溯源，明确在档案领域中人工智能的概念内涵，是解除档案人对人工智能误解的有效方法。

人工智能技术的应用是档案工作在经历计算机化、自动化、数字化、网络化发展后的信息化新阶段。所谓智能信息技术就是信息技术经历了上述发展阶段后更进一步的创新技术。目前，其相对成熟的部分正处于爆发式应用阶段，也就是说，经历前面每一阶段的技术积累，信息技术进入到人工智能技术阶段，前期的各种技术都是人工智能技术的支撑技术，是当前人工智能技术的组成之一或其底层技术。当底层技术都各自进入智能阶段后，完整的人工智能技术体系才真正形成，才能说人工智能技术全面进入到实际应用阶段。如图1所示是人工智能技术组成，即智能信息技术范畴。

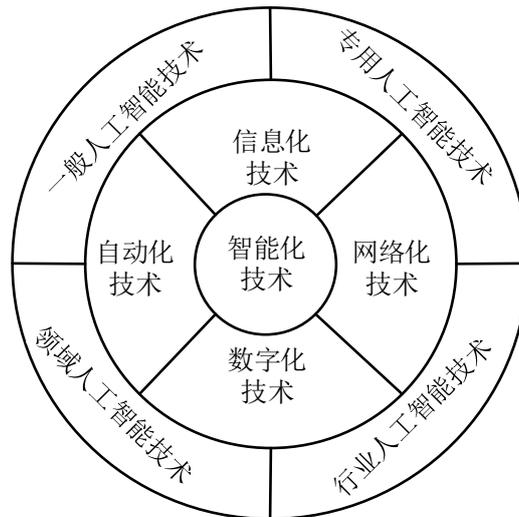


图1 人工智能技术组成

2.2 人工智能技术的档案内涵——智慧档案

目前，档案界尚未对智慧档案的定义达成一致，但笔者认为智慧档案的特征可从下列几方面来考查。

首先，智慧档案应该是成体系的人工智能技术运用于档案各环节后形成的全面智能化的档案生态。智慧档案是将档案工作、档案服务完全融入到智慧城市，成为智慧社会的有机组成部分后的一种档案形态。

其次，智慧档案是对数字档案功能的延展和提升，是对现代信息技术更深入、更广泛的应用，它将互联互通、协同工作、资源整合、价值发现、服务互动等人类所具有的主动、智能甚至是预知的思维充分展现出来。与数字档案相比，智能技术将在档案馆、实体档案、管理制度、从业者、信息基础设施、档案应用系统、档案信息服务、业务运作和档案责任者等方面实现全面信息化改造和升级[2]。

第三，智慧档案是新一代人工智能技术成熟后在档案领域的全方位应用。早期人工智能技术在档案工作中已有应用，1994年我国学者朱久兰就已经将人工智能技术运用于档案自动化管理[3]，当时所使用的人工智能技术与今天人们所熟知的大数据、云计算、物联网等技术有着很大的差别。用现在的眼光看，当时所使用的技术原始并较为复杂，智能效果也不明显，但却是现在人工智能技术的前身。而“2011年至今，人工智能技术突破了从“不能用、不好用”到“可以用”的技术瓶颈，人工智能发展进入爆发式增长的红利期”[1]。可以说，今天所说的人工智能技术是新一代的人工智能技术。

第四，智慧档案是档案信息化技术不断发展的新阶段。如图2所示，在档案领域先后经历计算机化、网络化、自动化、移动化、数字化、知识化、智能化过程后信息技术应用进入到智慧化阶段。从图2中可见，每一个发展阶段的实质都是对传统档案工作某种程度或某一方面的改进。这个“档案工作”不断改进的过程，也就是“档案工程”或“档案工作”智慧化的进程。与最初的计算机化相比，人工智能技术的运用，使从档案获取、档案管理、档案服务到档案利用的方式发生巨大变化，体现档案价值的档案信息载体也从“实体档案”、“档案数据”升华到“数字档案”、“档案知识”的智能阶段，即“智慧档案”的初级阶段。但档案工作全部环节的计算机化、自动化、数字化、知识化、智能化、智慧化仍然还有巨大的进步空间。其实，每次的档案信息化改造都赋予档案工作一些新的特性，这些新的档案特性也正是其所依赖的底层智能技术特性的抽象描述或特征概括。

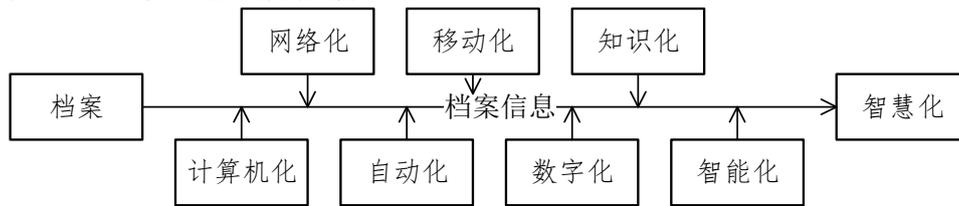


图2 智慧档案是档案信息化技术不断发展的新阶段

可见，在档案领域中信息技术应用通过几个阶段的发展，已经形成如图3所示的档案信息链体系结构。

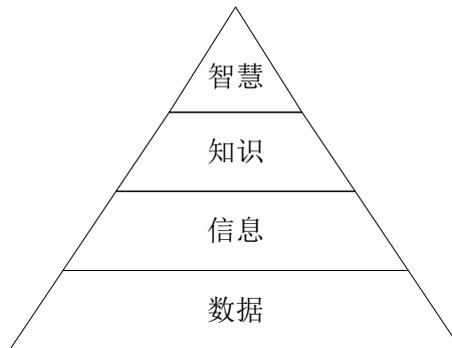


图3 档案信息链模式

第五，从智慧的内涵来看，人工智能技术仍然处于智能阶段，而非智慧阶段。当前，智慧档案工作采用的主要技术是信息感知技术、机器学习技术。所谓人工智能技术在档案领域的应用，笔者理解为主要是指发展相对成熟的专用人工智能技术在档案领域的应用。从所使用的技术角度来说，有计算机视觉技术、语音识别技术、自然语言理解、机器翻译技术等，而不是指通用人工智能技术。从应用技术在档案领域适用性来说，主要是档案工作各环节，比如：档案库房、档案数字信息采集、档案管理、档案应用等。因此，当前人工智能技术仍然处于智能阶段或说智慧的初级阶段。

2.3 智慧档案系统

在档案工作中运用人工智能技术的历程如图4所示。不难看出，在智慧档案中人工智能技术不是指一项技术，而是指信息处理各环节的智能化技术，而智慧系统就是当该项工作的信息处理各环节全部由智能技术实现其功能时的系统，换句话说就是信息处理各环节的功能由智能技术实现的系统，即由各专用智能技术组合而成，实现特定功能的系统，才称为智慧档案系统。因此，智慧档案的概念内涵应该理解为档案各环节普遍采用数字化、智能信息技术实现其功能的现代档案生态^[4]。



图4 当前成熟的专用人工智能技术与其应用场景图

3. 智能信息技术在档案领域中的国内外应用现状

国内外档案界，在智能技术迅速发展的背景下，正在探索档案与智能技术的融合应用。在法德NULB^[6]项目中，使用了自动语音识别技术和机器学习技术归档无文字的语言；在意大利IPSA^[7]项目中，将传统文化收藏品数字化后，研究了人们对数字化传统文化收藏品的参与度；葡萄牙的波尔图大学通过对归档门诊记录的分析，研究了产科信息系统VCObsCare在医疗实践中的有效性^[8]；2018年希腊比雷埃夫斯大学提出的SPARTAN框架采用了大数据技术和语义集成技术有效集成和处理了不同来源时空的大数据^[9]；2017年加拿大的达尔豪斯大学采用先进机器学习技术、图像数字化技术、数据集成和可视化技术研发了利用数字图像档案和社会媒体资源的社会影响评价工具cultromics^[10]；2017 印度管理学院和新加坡国立大学，基于183个国家的档案数据，研究了政府电子数据开放和电子政务成熟度的关系，以及政府在电子政务成熟度和技术组织环境之间的中间人角色^[11]；2014年土耳其哈斯特帕大学研究了数字图书馆的可用性测试^[12]；2018是法国数字档案的转换点^[13]；2018西班牙的阿威罗大学开发了可逆去识别的医学图像档案的可控搜索框架^[14]；2018印度阿卡龙雅理工学院提出基于机器学习算法的交互式文本挖掘框架来实时识别药物名称^[15]。从以上文献不难看出国外在档案数字化、档案信息技术的运用、档案人工智能技术的运用方面近年来呈现如下趋势：1) 全球档案领域表现出积极发展的趋势，但是显然还不够成熟，处于初步探索阶段；2) 在发展中国家和地区，比如印度和中国发展态势比传统发达国家更积极；3) 当前热门的专用人工智能技术应用于档案领域还较少见，发展潜力还很大。

国内现状：我国的智慧档案建设已经起步，但主要还处在理论研究阶段，其中学术研究是主力，应用研究相对较少；实践层面的工作呈现出逐步探索和逐年增加态势。首先从档案智能信息装备的发展情况来看，国内文献在档案数字装备研发方面的报道较少；从产品研发来看，对档案管理的应用装备研发的企业和研究机构还较少，行业标准还未出现；从实际应用来看，可选择的成套产品及配套产品种类还远远不够丰富，还远不能满足信息化、数字化、数据化档案管理与档案利用的需求。其次从应用软件功能和性能来看，人工智能技术对档案利用的支持还远远未达到人们所期望的状态。可以说，智能装备与人工智能技术在档案工作中的应用才刚刚开始。主要的工作有：上海档案馆的远程利用服务体系，实现智慧档案馆智

慧服务的主要内容是融入智慧城市的档案服务^[16]；青岛市建立档案智慧收集平台，采用物联网和射频技术实现档案实体电子标签化，并据此建立档案智慧管理平台，实现识别、定位、跟踪、监控和管理的智能化，进一步引入知识管理、数据仓库、三网融合、移动WEB技术建立档案信息智慧泛在的智慧档案服务平台；采用虚拟化技术构建档案信息智慧保护平台，实现智能备份、数据安全的智慧检测服务；构建机关和企业档案智慧监督平台^[16]；韶关市国土资源档案馆档案智能化管理建设平台，借助无线射频识别（RFID）技术，有效建立统一标准的档案信息化管理数据库，对实体档案进行智能化安全管理，为档案管理的“收、管、用”三大业务环节提供便捷的服务通道，取得一定成效^[17]。第三，从理论成果来看，虽然提出一些智能应用框架，但各框架间还看不出有统一的顶层机制设计，为各种框架间埋下了屏障问题。典型的智能应用框架如高效档案系统智能管理中^[18]，以档案数据与数字资源为主要内容，其在一体化管理系统支撑下，采用基于网络环境的信息系统对档案资源进行精细管理，通过在线服务系统和到馆服务系统，向用户提供不断更新的档案信息服务，主要流程为：接收→鉴定→存储→管理→安全维护→检索分类→档案编研。该智慧档案体系结构完整、全面，自成体系，但潜在地与其他体系结构不相兼容，容易形成信息孤岛和共享障碍。因此，保障档案资源可感知、能互联、优服务、强整合、勤利用的工具亟待研究和开发^[16]。

综上所述，目前大多数档案部门在“智慧档案”或“智慧档案馆”的探索和建设方面都还处于概念阶段或是理论探讨阶段，唯有以青岛市智慧档案馆为代表的少数几个单位已经实现初步的智慧档案馆建设。尽管其仍在探索和建设，很多内容、功能、实际效果等还不确定，但它的启动和建设已经标志着我国智慧档案馆的兴起，这也将为我国智慧档案馆的理论研究提供典型案例，并为我国全面建设智慧档案馆提供参考和借鉴。在“大数据技术”和“知识工程”理念驱动下，档案信息化的条件不断完善，但档案社会化媒体信息资源整合开发的研究鲜有涉及，有关技术手段的研究明显缺乏^[5]。在高校档案数字化、留学生档案智能化和智慧化方面的专项研究也表现出明显不足。

4. 智能信息技术在档案领域中的应用展望

人工智能技术向社会生活的各领域渗透是信息技术发展的必然，“智慧档案馆”概念的提出正是适应这种发展需求的表现。随着智慧城市的建设，作为其一部分的智慧档案馆必然要完成其自身的智慧化，而档案智慧化的实现应该是进行时态，仍处于发展过程初始阶段，仍将伴随着人工智能技术的发展而进步完善。之所以这样说，是因为智能是智慧的基础，目前我们所运用的技术，实现的只能说是低级智能或说初级的智慧^[1]，那种由人的思维而控制的智能是未来的高级智能，而大家之前所说的智能并没有体现出初级智能的内涵，因而将智能简单理解为信息系统+物联网就是智能是一种误解。因此，如何在数字化基础上首先实现智能化，进而实现智慧化、如何融入智慧城市中、如何提供更好的智慧服务，都是需要档案人展开探索的未知领域。

基于对智能技术与智慧技术的上述理解，我们认为档案工作当前应该以智能技术的普及为先，因为智能技术是智慧技术应用的前提和物理基础。因此，从“档案工程”角度看，智能基础设施的建设在前，智慧技术的运用在后。但这绝不是忽视智慧技术的应用研发，从技术发展的角度来看，智慧技术的运用是档案工作服务社会的下一代接口，没有应用技术的提前研发和储备将在后期档案服务工作中陷入被动。

总之，这些新的概念的演变过程，体现出信息技术发展对档案事业的影响、融合、分工和支撑作用，体现了档案人与时代接轨、与时俱进的精神风貌。但是，新技术的引入也给档案人带来了许多新问题，特别是面对迅速发展、扑面而来的新技术，容易使人感到困惑、无助与迷茫。

5. 智能信息技术在智慧档案应用中可能的创新点

智慧，笔者将其理解为智和慧两部分，智即智能技术，慧即聪慧技术。智能更多偏向物联网、嵌入式等偏硬件的基础设施技术，聪慧则更多指大数据和深度学习等偏软件的技术。基于智能技术与聪慧技术的理解，智慧档案中所说的人工智能技术是档案数据层、应用层、基础功能层中一系列技术的总称，并非单指一项技术。因此，目前可预见的突破点可能会在下面几个方面：

1) 基于大数据和深度学习的数据分析能力的创新研发会使档案的利用水平更进一步。

2) 基于物联网技术、嵌入式技术的智能传感器等应用装备的研发，使档案获取、管理、服务、利用具备自适应能力。基于华为短距传输物联网硬件技术的智能档案装备研发有着可预见的前景。

3) 目前各业务系统的格式不统一，是难以实现自动归档的主要问题。减轻档案资源采集的劳动强度及无纸化办公理念下的档案资源数字化，电子印章、数字印章、加密印章的研发是解决档案资源自助采集的前提。档案资源自助采集或自动归档的前提又是统一的底层“数据格式”和上层“多媒体表示标准”。因此，自助档案资源数字化及新媒体档案资源标准研发也是急需突破的重要方向。

6. 结束语

通过以上分析，不难看出档案工作是一个系统工程。因此，智慧档案是一个伴随信息技术进步而不断演进的过程。信息技术虽然进入人工智能技术阶段，但当前人工智能技术正处在由智能技术向智慧技术演进的过程中，即只是在某些个点上有了突破、具备了实用性。而从技术体系完整性来看，尚未形成完整的人工智能技术体系，还不足以支持档案工作全部环节的智能化和智慧化。因此，智慧档案的真正实现取决于人工智能技术体系完整性的进程，智慧档案工程必将是一个长期的演化与递进过程。

References

- [1] Tieniu TAN, Zhenan SUN, and Zhaoxiang ZHANG, Artificial intelligence: angel or devil?, SCIENTIA SINICA Informationis, vol.48(9), pp. 1257-1263, 2018.
- [2] Sixin XUE, Jijun YAN, and Fenhua Hu, Explore and Analyse of Modern Management Development in Universities Archives , China Archives, vol.(5),pp. 61-63 , 2018.
- [3] Jiulan ZHU, Smart Retrieval of Information of Archives, Archives Science Bulletin, vol.(4), pp. 39, 57-58, 1994.
- [4] Lan WANG, From Letter-writing to Cloud -- Thoughts and Inspirations of Archives Ecological Changes, The second academic seminar on informatization and intelligence archives, 2018.
- [5] WANG Lancheng and HUANG yongqin, Research on the Mining and Utilization of Archives SocialMedia Information under the Background of Big Data , Archives Science Study, vol(6), pp. 73-76, 2015.
- [6] Adda G, Stüker S, and Adda-decker M, Breaking the Unwritten Language Barrier : the Bulb Project, Procedia Computer Science, vol.81, pp. 8-14, 2016.
- [7] Agosti M, Ferro N, and Orio N, Cultura Outcomes for Improving the User's Engagement with Cultural Heritage Collections, Procedia Computer Science, vol.38, pp. 34-39, 2014.
- [8] Barrote A, Silva P, and Gonçalves F, Obstetric Information System: Effectiveness in Health Care Practice, Procedia Technology, vol.16, pp. 1411-1416, 2014.

- [9] Santipantakis GM, Glenis A, and Patroumpas K, Spartan: Semantic Integration of Big Spatio-temporal Data From Streaming and Archival Sources, *Future Generation Computer Systems*, 2018.
- [10] Sherren K, Parkins JR, and Smit M, Digital Archives, Big Data and Image-based Culturomics for Social Impact Assessment: Opportunities and Challenges, *Environmental Impact Assessment Review*, vol.67, pp. 23-30, 2017.
- [11] Krishnan S, Teo TSH, and Lymm J, Determinants of Electronic Participation and Electronic Government Maturity: Insights From Cross-country Data, *International Journal of Information Management*, vol.37(4), pp. 297-312, 2017.
- [12] Dalkıran Ö, Aker İ, and Öztemiz S, Usability Testing of Digital Libraries: the Experience of Eprints, *Procedia - Social and Behavioral Sciences*, vol.147, pp. 535-543, 2014.
- [13] Chabrol B, 2018, a Turning Point for the Archives De Pédiatrie, *Archives De Pédiatrie*, vol.25(1), pp. 1-2, 2018.
- [14] Silva JM, Pinho E, and Monteiro E, Controlled Searching in Reversibly De-identified Medical Imaging Archives, *Journal of Biomedical Informatics*, vol.77, pp. 81-90, 2018.
- [15] Chukwuocha C, Mathu T, and Raimond K, Design of an Interactive Biomedical Text Mining Framework to Recognize Real-time Drug Entities Using Machine Learning Algorithms, *Procedia Computer Science*, vol.143, pp. 181-188, 2018.
- [16] Zhiyong YANG and Feng ZHOU, Analyze the Rise and Future Development of Smart Archives, *Archives Science Bulletin*, vol.(4), pp. 45-50, 2015.
- [17] Shunfang Li, Overview of Intelligent Management Practice about Land and Resources Archives in the Era of Big Data, *Archives Science Study*, vol.(S2), pp. 50-52, 2017.
- [18] Xuelian GAO, Design and Implementation of Environmental Monitoring and Management System for Archives, *Archives Science Study*. vol. (3) , pp. 50-52 ,2017.
- [19] BAI Wenlin and AN Xiaomi, Research Status, Hot Fields and Prospects of Archival Science in 2014. *Archives Science Study*, vol.(3), pp. 12-15, 2016.