

基于CiteSpace的信息共享背景下档案数字化建设应用研究

罗敏超, 金一然, 石若瑶, 魏 强*

江汉大学, 湖北 武汉

收稿日期: 2022年9月14日; 录用日期: 2022年10月17日; 发布日期: 2022年10月24日

摘 要

近年来, 共建、共享理念的提出无疑给城市智慧化的发展带来了新元素, 而将其运用到公共文化服务领域能够提升国家文化软实力。因此依托“互联网 + 信息传输技术”、大数据及云计算技术等现代技术建设数字化档案管理体系已经成为了重点研究方向。本文以知网作为数据来源, 以档案数字化、共享作为检索关键词, 选取了2011年到2021年10年间的中文相关文献606篇, 并结合传统的文献计量方法, 从多个视角来分析我国的档案数字化建设的研究现状、趋势与热点等, 统筹安排档案数字化推进工作, 推动数字档案资源的整合与共享, 努力为医疗、媒体、旅游、教育、银行等企事业单位提供一站式服务。

关键词

共享, 档案数字化, 档案智慧化

Research on the Application of Archives Digital Construction under the Background of Information Sharing Based on CiteSpace

Minchao Luo, Yiran Jin, Ruoyao Shi, Qiang Wei*

Jiangnan University, Wuhan Hubei

Received: Sep. 14th, 2022; accepted: Oct. 17th, 2022; published: Oct. 24th, 2022

Abstract

In recent years, the concept of co-construction and sharing has undoubtedly brought new ele-

*通讯作者。

ments to the development of urban intelligence, and applying it to the field of public cultural services can enhance the country's cultural soft power. Therefore, relying on modern technologies such as "Internet + information transmission technology", big data and cloud computing technology to build a digital archives management system has become a key research direction. This paper uses CNKI as the data source and archives digitization and sharing as the retrieval keywords, selects 606 Chinese-related documents from 2011 to 2021, and combines traditional bibliometric methods to analyze my country's data from multiple perspectives. The research status, trends and hotspots of archives digitization construction, make overall arrangements for the promotion of archives digitization, promote the integration and sharing of digital archives resources, and strive to provide one-stop services for medical, media, tourism, education, banking and other enterprises and institutions.

Keywords

Shared, File Informatization, Archive Digitization

Copyright © 2022 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

“档案数字化是指利用计算机等技术手段将经过鉴选的、原本存储于传统载体上的档案资料进行数字化转换” [1]。其内容包含资料的接受、归类、编制、检索以及调取等，这构成了档案系统完整的工作流程。2002年，《全国档案信息化建设实施纲要》提出要加快传统档案的数字化工作[2]。有关部门在2006年首次提出，要通过现代化技术的使用来优化当前的档案管理模式，全面实现数字化管理[3]。2011年，《全国档案事业发展“十二五”规划》要求“加快推进传统载体档案数字化” [4]。2020年10月第三届数字中国建设峰会召开，会议讨论了如何将数字化思想融合进生活的多个领域，其内容涵盖了政府、经济、健康、生态、智慧城市与社会、大数据与区块链、智能制造与物联网等多个方面，让观众深度领略了飞速发展的数字中国。陆续出台的相关政策方针，充分体现了我国对档案数字化转型工作推进的重视。

在党的第十九次全国代表大会上习近平总书记提出了“打造共建共治共享的社会治理格局”，这一理念的提出进一步鼓励和吸纳了社会力量的有序参与并实现公共利益的最大化[5]。在人工智能、大数据等技术快速发展的新时代背景下，档案资源的利用与服务也产生了革命性的变化，它作为新兴信息载体越来越多的服务于社会和生活等各个领域，因此如何更有效的提高档案资源建设与应用的水平、如何更好的发挥部门与部门、企事业与企事业合作的联动力量、如何最大限度的发挥档案资源共建共享以有效解决档案管理中的信息孤岛[6]，成为一项有待解决的问题。尤其要重视公共文化机构的建设，例如图书馆、档案馆、博物馆，这有助于实现我国建设文化强国的战略目标[7]。

为了客观的总结我国飞速发展的十年间，在档案数字化管理、转型、研究等方面的发展情况，本文以中国知网(CNKI)为收集数据的对象，对相关期刊文献采用可视化方法，结合文献分析对前人所做的档案数字化的共享方面的研究进行梳理，同时提出新的观点和想法，为档案数字化的应用起到推动力量，将其充分与社会生活相结合，切实满足企事业单位以及个人的需要。

2. 数据来源与处理

通过查阅相关文献，早在2014年已有研究者采用了CiteSpace数据分析软件对我国档案数字化的研

究热点进行了梳理,但梳理范围过于宽泛,并没有着重突出档案数字化转型工作最重要的方面即数据共享,因此本文以中国知网(CNKI)作为收集数据的来源,以“档案数字化”为主题,并将共享作为关键的检索词检索相关文献。检索条件:(主题 = “档案数字化”or 题名 = “档案数字化”)AND (主题 = “共享”or 题名 = “共享”)AND (发表时间 Between (“2011-01-01”, “2021-12-31”));检索范围为:中文文献;其中学术期刊 606 篇,导入 CiteSpace 后有效转换记录 606 篇,设置年度切片默认时长为 1 年,阈值为前 50,然后对关键词等进行分析、对知识图谱进行处理,以此将档案数字化工作中数据共享的部分作为主要探讨内容,并梳理更新的相关内容。

3. 研究工具与方法

本文采用美国德雷塞尔大学信息科学与技术学院陈超美博士所开发的软件 CiteSpace [8],该软件可根据数据分析的需求对学术文献数据进行量化,形成可视化知识图谱。根据需要,图谱中会显示出关键词、聚类、聚类之间的联系以及随时间的演变。这是近年来在科学计量学、知识计量学领域中新兴的一种定量分析法,它能够通过一系列可视化图谱直观的揭示其相互关系,理清该研究领域的发展脉络。本文主要采用 CiteSpace5.8.R3 软件对“档案数字化”和“共享”两个关键词进行可视化和知识图谱分析,选择关键词为分析节点,聚类后再绘制时线图,并对凸显词进行分析,然后从分析出的研究热点、研究前沿等入手,采用文献分析法,对我国近 10 年来有关档案数字化理论以及实践层面的相关研究进行扩展延伸,提供前沿热点下的档案数字化建设的建议。

4. 研究现状

文献来源与发文量上的分布显示了该研究领域在宏观上的总体状况,其文献来源可代表该研究领域所关注的角度。发文量可在一定程度上代表研究领域的热度,发展规模及速度可作为一个重要的衡量指标。

如图 1 所示,经统计分析,文献主要的学科分布集中在档案与博物馆方面,占比 52.46%,其次为计算机科学与计算机应用,占比 21.47%,即包含共享成分的成果主要通过计算机等现代科技软件实现。发文量排前五的期刊分别是《办公室业务》59 篇、《兰台内外》53 篇、《黑龙江档案》47 篇、《城建档案》和《兰台世界》分别有 46 篇和 35 篇。经研究发现,这几类期刊主要报道在数字档案、科学管理、智慧档案局馆等领域的理论与创新研究,对档案数字化这一研究领域具有一定的学术影响力。

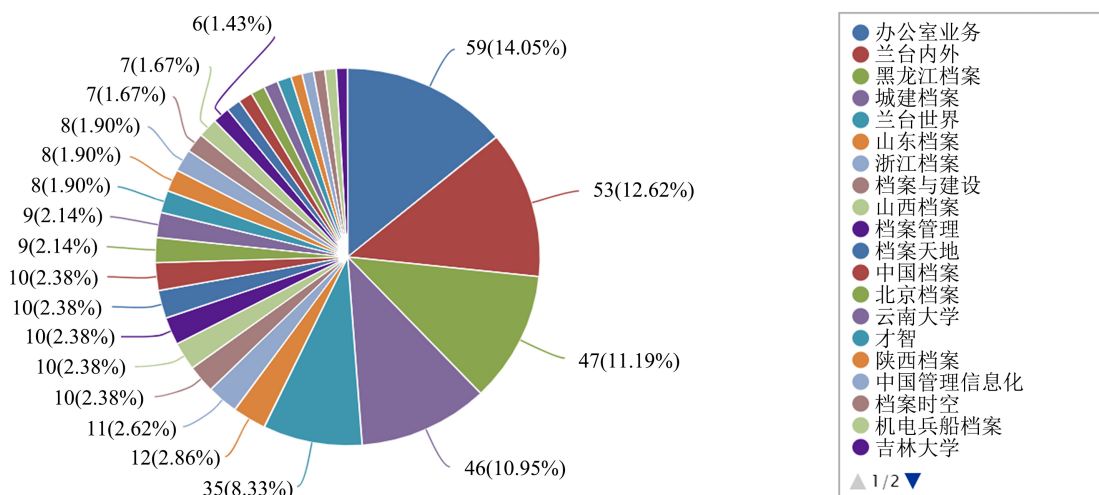


Figure 1. Distribution of sources of literature and journals
图 1. 文献期刊来源的分布

如图 2 所示, 经统计分析, “档案数字化 + 共享”类文章, 在 2011 年至 2021 年间一直处于稳步增长的状态, 年均发表文章 40 篇左右; 2018 年至 2019 年属于快速发展阶段, 几乎达到年均发表 100 篇的数量; 2020 年略波动起伏, 直至 2021 年稳步回升, 发文量在 93 篇左右。经研究发现, 中国国家档案局在 2000 年至 2016 年间先后制定了 4 个全国档案事业发展五年计划[9], 这是导致在相关领域研究持续增长的一个重要原因, 并且在 2018 年, 我国在两会上提出要落实新的科技政策, 把握新一轮的科技革命, 因此在全国范围内铺开机构改革工作, 档案相关部门加强公共服务能力的改革工作也开始正式启动。

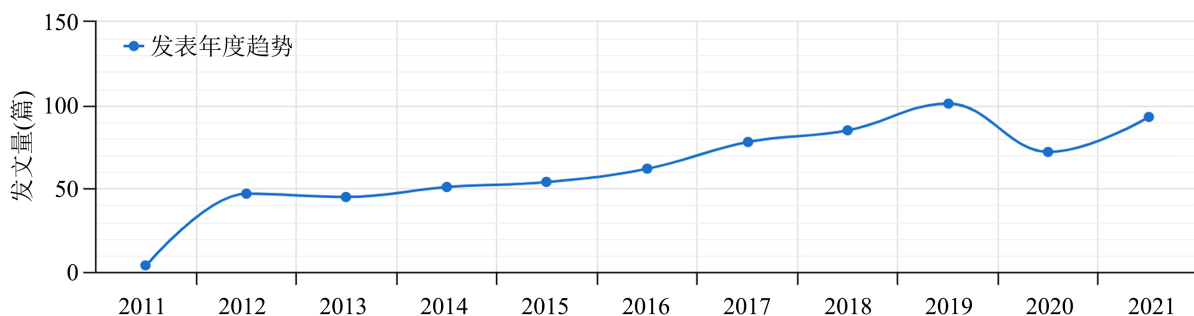


Figure 2. Annual distribution of the number of articles

图 2. 文献数量的年度分布

5. 基于 CiteSpace 档案数字化的知识图谱分析

5.1. 关键词共现分析

关键词是一篇文献主题的凝练与概括[10], 通过关键词能够快速把握文献的中心思想。关键词在多篇文献中共现以及不同时期下关键词共现的变化, 能够说明该研究领域的研究热点以及发展方向, 深挖这一大类文献有助于我们抓住该领域的关注焦点。在 CiteSpace 软件中, 频率和中心度是衡量关键词重要指标。频率越高, 标签越大; 中心度值越大, 节点越大[11]。本文选定关键词为网络节点类型, 时间跨度设置为 2011~2021 年, 从绘制的档案数字化研究关键词共现知识图谱可见, 共有 295 个关键词节点和 548 条节点间的连线, 且网络密度为 0.0126 (如图 3)。对排名前 20 的关键词共现频率进行了排序(如表 1), 其中, 频率最高的是“数字化”, 其次分别是“档案管理、信息化、档案、高校、管理、高校档案、数字档案、信息共享”等, 这部分关键词词频较高, 这说明了档案数字化建设是实现数字资源共享的前提。

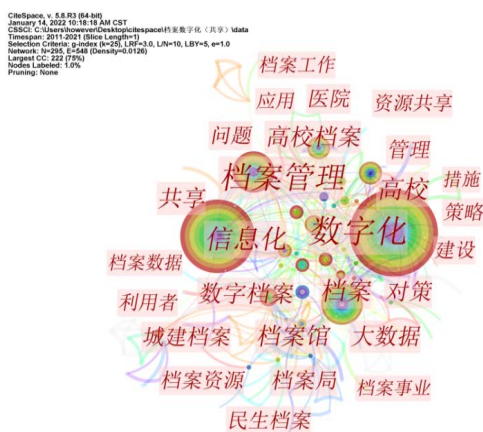


Figure 3. Archival digitized keyword co-occurrence map

图 3. 档案数字化关键词共现图谱

Table 1. Statistics of word frequency and centrality of the top 20 keywords in archival digitization research
表 1. 档案数字化研究前 20 位关键词词频、中心度统计

序号	频率	中心度	年份	关键词	序号	频率	中心度	年份	关键词
1	128	0.4	2011	数字化	11	16	0.02	2012	信息共享
2	101	0.33	2011	档案管理	12	14	0.1	2013	档案馆
3	42	0.13	2011	信息化	13	14	0.05	2012	对策
4	38	0.05	2011	档案	14	13	0.06	2015	民生档案
5	28	0.07	2011	高校	15	13	0.03	2011	档案信息
6	23	0.02	2011	管理	16	13	0.1	2018	档案数据
7	23	0.05	2011	高校档案	17	12	0.02	2014	医院
8	20	0.07	2012	数字档案	18	12	0.01	2018	问题
9	16	0.07	2016	大数据	19	11	0.03	2015	应用
10	16	0.07	2012	城建档案	20	11	0.08	2012	共享

结合相关文献进行深挖,在这些关键词中,管理占据一个突出地位,这表明了建立数字化共享服务需要完善统一的组织体系来对档案管理网络化构建以及资源共享进行规划与协调[12]。强调要加强顶层设计、实现区域发展平衡,这可依托政府现有的对接政务的已有资源来构建格式规范、数据集中、可共享的数字资源库,以突破时空和地域界限,达到信息共享的目的[13]。另外信息化这一关键词强调了网络发展的速度已经达到了可以为个人或者企业提供个性服务的阶段,因此档案服务也紧随信息化的发展趋势,建立档案共享中心(AI 中心、业务中心、创新中心、协同中心),通过统一接口为业务系统提供数据服务支撑,应用智能语音等提升档案业务服务智慧化[14]。对信息共享的解读主要体现在政务部门的档案资源共享如何应用于教育、医疗、新闻媒体、旅游行业、城建、企业等行业内外。因此如何充分利用现有数字化档案资源共个人使用,行业内外开发也是目前的热点,此研究有利于达成双向的互动和需求的满足。

5.2. 关键词聚类分析

从关键词共现分析中提取聚类命名术语可得到可视化的关键词聚类视图(如图 4)。图谱的绘制效果可用模块值(Q 值)和平均轮廓值(S 值)两个指标来衡量[15]。一般而言,当 $Q > 0.3$ 时,表明划分出来的网络结构是显著的;当 $S > 0.5$ 时,一般被认为聚类是合理的;当 $S > 0.7$ 时,聚类是高效率且令人信服的[16]。如图五所示,聚类结果统计出 10 个聚类(Q 值为 0.5428、S 值为 0.8417),说明绘制的聚类图谱效果较好,其中最主要的聚类包括:“数字化、对策、档案管理、城建档案、高校档案、信息化、利用者、档案利用、档案工作、电子档案”。

结合相关文献进行分析表明,办公自动化、网络信息共享等技术在政府和企事业单位的普遍应用,电子政务、数字城市等工程逐步实施[17]是目前在档案数字化领域研究的热点,是加快档案现代化管理的迫切要求,也符合社会发展的实际需要。无论是政府、企事业、民生这类收集能全民信息的档案,还是

教育、城建、医疗这类能促进本行业发展的档案,在推进档案数字化工作时都强调要分别在技术、系统、人员、意识等方面做出突破。在技术方面,如何建立完备的信息云数据库、建立智能的数据处理优化系统、建立完整的终端应用体系[18]来实现技术上的突破是有待解决的问题。在系统方面,建立集中统一的数字档案系统以及统一标准的技术路线是建立数字化管理平台所必须的考虑的。各机构部门也要突破传统的媒介管理体系、地域局限的限制,形成档案数字化成果之间的关联与共享。然而提升档案相关人员的数字化意识也是档案数字化建设的目标之一,在专业档案人员掌握丰富的档案管理工作理论与实践的基础之上,提升其信息技术水平,将“软知识”和“硬数据”相结合,变成复合型技术人才。

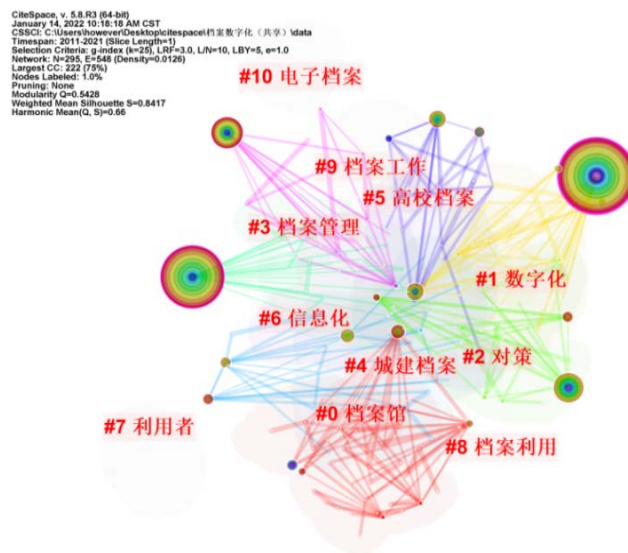


Figure 4. Keyword cluster analysis visualization map
图 4. 关键词聚类分析可视化图谱

5.3. 关键词时线图分析

时线图是一种侧重于从时间维度上表示知识演进的视图,它可清晰展示出关键词的更新和相互影响[19],图中节点代表关键词,连线表示关键词节点之间关系,横坐标表示时间,纵坐标为关键词聚类标签。因此,本文利用 CiteSpace 进行时线可视化分析以探讨我国信息共享背景下档案数字化研究领域的前沿。依托聚类生成的结果截取前 7 个影响较大的聚类,其余 3 类聚类影响力小,本文暂不采用,由此生成的关键词共现时线图(如图 5)。

结合相关文献分析可知,该领域的研究热点与时俱进,不断更新。早在 2008 年之前,我国关于档案的相关研究主要是以档案工作为主,将其关注点放在档案的管理规范、图书档案的利用,后来档案馆开始追求新的管理模式和利用模式,将其档案进行综合,推行档案共享服务、拓宽档案的存储空间,以此来达到优越性的档案实施和管理模式,这也体现了档案管理的演进与发展。2010 年进入一个新阶段出现了档案数字化趋势。在这一时期,数字化工作统筹协调,遵循“资源建设,格式规范,整体协调,安全落实,经费保障”等原则。对高校、民生、医疗档案等相对较早时期的各部门档案在档案应用性方面进行研究,显示了将传统档案慢慢转变到档案服务的演变过程,重点考究有哪些部门可以将其转化,慢慢的随着科技的发展开始出现医疗档案的创新研究。在 2016 年以后,开始强调共建共享,共享中心等以数据和资源形式开展数字化工作,这也表明了我国档案工作在大数据时代下不断推陈出新,也为后续研究共建共享平台下的档案数字化提供了方向。

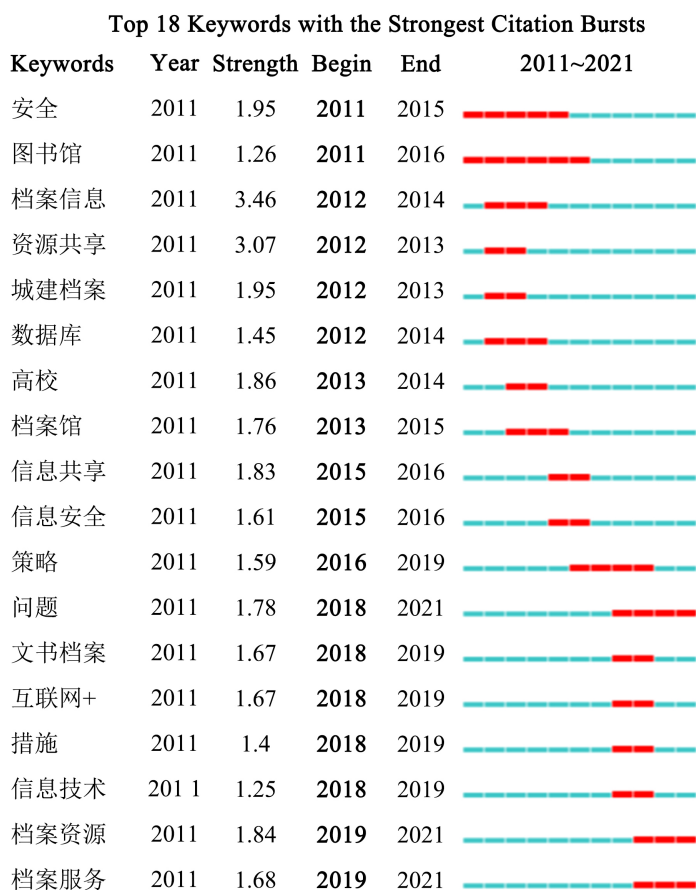


Figure 6. Keyword emergence analysis of archives digitization

图 6. 档案数字化关键词突现分析

6. 档案数字化相关研究展望

档案数字化建设的研究进程随着信息技术的发展已经开展了 20 余年,结合关键词、聚类以及凸现词等综合分析来看,未来,相关研究者还会持续在建设和推进档案数字化的基础性工作方面做出努力,加速推进数字化共享的应用。而做到这一点,离不开更高水平的科学技术的支撑,可在其基础上将研究精力投入到重非数字档案的数字化、建立档案数字化专题类数据库、扩大档案数字化共享服务范围等方面。首先,就是要保障档案资源的齐全、完整,并有效管理和便于利用,重点开展照片和音像等这类难度相对较高的档案工作,使其档案信息得到最大化利用。其次,相关部门在引进先进设备的基础之上,对现有的海量档案资源进行分析和重组,按照某种联系或不同的分类标准归类、排序,建立专题类数据以加强数据资源之间的关联程度。最后档案数字化的共享服务的建设应当充分发挥各主体在社会界上发挥的作用,以下几个方面是推行数字化共享的前沿方向。

1) 政务部门的档案资源共享。各省市每年所统计的各项指标数据、民生档案、发展计划等大部分资源主要掌握在政府等相关部门,因此政府借助信息技术优化当前档案数字化建设的架构体系,让数字化的档案变成活水源头在各行各业充分享受档案数字化资源共享所带来的便利条件。

2) 企业、行业内部的档案数字化的应用。医疗、媒体、旅游、高校教育、银行、城建等充分收集内部资料、数据,依托互联网等信息技术实现资源的数字化分享与利用,为教师和学生提供专门的学业指导,为医院以及求医者提供广泛的求医路径、治疗方案,也可以通过档案的数字化等实现为定向人群推

送有力旅游信息资源。

3) 提供便民服务系统。目前“一站式服务”符合政府、企业以及个人所推崇的简化、便捷化的理念。因此继续拓展跨部门和跨省便民档案利用服务的业务范围、开发一款便民服务的 app 和小程序将所有资源进行整合一键式查询并调取相关信息,是扩展数字资源的共享的有效途径。档案相关工作人员、研究者应当抓住当下研究热点,继续推行档案数字化工作,有助于档案管理服务智慧化水平全面提升。

参考文献

- [1] 屠跃明, 翟瑶. 档案数字化的元数据研究[J]. 兰台世界, 2012(14): 60-61.
- [2] 屠跃明, 钱毅, 李洪. 关于档案数字化质量控制体系的研究[C]//创新: 档案与文化强国建设——2014 年档案事业发展研究报告集. 中国档案学会. 中国文史出版社, 2014: 114-134.
- [3] 王悦. 数字化档案管理系统在医院中的应用研究[J]. 财经界, 2021(20): 59-60.
- [4] 严雪. 大数据背景下科研档案文献数字化趋势[J]. 兰台世界, 2018(8): 81-84.
<https://doi.org/10.16565/j.cnki.1006-7744.2018.08.22>
- [5] 李明悦. 共建共治共享社会治理理念研究[D]: [硕士学位论文]. 哈尔滨: 东北农业大学, 2019.
- [6] 李若池. 数字化科学教育资源共建共享机制研究[D]: [硕士学位论文]. 武汉: 华中师范大学, 2021.
- [7] 张丹. 综合档案馆数字档案资源建设存在的问题与对策研究[D]: [硕士学位论文]. 郑州: 郑州大学, 2017.
- [8] 滕金润, 蔡云. 基于 CiteSpace 的国内汽车热管理文献可视化分析[C]//四川省第十五届汽车学术年会论文集. [出版者不详], 2021: 208-213. <https://doi.org/10.26914/c.cnkihy.2021.023443>
- [9] 李明华. 中国的数字档案资源建设[J]. 中国档案, 2016(10): 14-15.
- [10] 邵永斌. 基于 CiteSpaceV 的人事档案系统与高校档案数字化可视分析研究[J]. 兰台世界, 2019(7): 51-54.
- [11] 谢瑶瑶, 乔硕功, 向禹. 基于 CITESPACE 的档案工作(2008-2017)研究热点可视化分析[J]. 资源信息与工程, 2018, 33(1): 188-191.
- [12] 王春雨. 数字化时代档案管理技术创新发展措施研究[J]. 兰台内外, 2021(27): 40-41.
- [13] 尹哲. 大数据服务档案文化建设的问题与对策[J]. 北京档案, 2018(12): 21-23.
- [14] 加快档案资源数字化转型, 提升档案数字治理水平[J]. 浙江档案, 2020(11): 9-12.
- [15] 曾小桥, 卢东, 贺碧玉. 我国旅游共享经济研究评述——基于 CiteSpace 的文献可视化分析[J]. 资源开发与市场, 2018, 34(8): 1179-1184.
- [16] 张博, 康奥. 国内地质公园研究演化趋势与热点分析——基于知识图谱视角[J]. 地质论评, 2021, 67(2): 557-569.
<https://doi.org/10.16509/j.georeview.2021.021>
- [17] 王学平. 浅议我国档案数字化建设实践与发展策略[J]. 档案学通讯, 2011(6): 54-57.
- [18] 胡雅静, 周林仙. 基于 CiteSpace 的干部人事档案信息化研究综述[J]. 办公室业务, 2021(15): 75-77.
- [19] 袁帆, 梁孟华. 中外电子健康档案管理研究热点与前沿趋势——基于 VOSviewer 与 CiteSpace 可视化分析[J]. 山西档案, 2021(3): 151-162+150.
- [20] 赵帆, 刘娅. 基于文献计量视角的中国与东盟科技合作状况研究[J]. 世界科技研究与发展, 2020, 42(4): 376-387.
- [21] 韩玉玲. 大数据时代档案管理创新路径分析[J]. 山东档案, 2020(6): 72-73.