

大数据挖掘技术在档案管理中的运用

张林

济南东城实验高级中学 山东 济南 3456

[摘要]进入二十一世纪以来,随着社会发展,我国的科学水平得到进步,推动档案管理逐渐朝着信息化的方向发展。特别是档案数据挖掘技术的出现,其依托于大数据技术,实现了档案管理的精细化,满足了用户与档案资源的双向控制,促进了新时期档案管理质量与效率的双提升。在档案管理中实施档案数据挖掘,需要根据实际情况以及发展需求进行,在提升信息处理质量与效率的同时,实现进一步发展。本文就大数据技术概述、档案数据挖掘技术、档案数据挖掘现状、大数据技术背景下的档案数据挖掘等方面展开论述与分析。

[关键词]数据挖掘技术;技术应用;档案管理系统

[DOI] 10.12252/j.issn.2096-627X.2021.09.1655

引言

随着社会发展不断深入,无论是数据信息的规模还是种类均呈现出高度的多样化。由此也进一步加大了档案管理工作强度,面对庞大的档案管理诉求,如若单纯地沿用传统的档案管理模式势必无法达到实际诉求。计算机技术的引入和使用,可以很好地代替传统档案管理模式,大大提升档案数据等的安全性以及有效性,此外也可以更为客观、科学地提升档案管理水平。故以大数据为环境等新型信息挖掘技术等在现代档案管理中的运用有较高的现实意义。

1 数据挖掘技术概述

数据挖掘技术是指借助一定的手段对已知数据加以分析、整合,从而获取到其内在的信息的一种技术。数据挖掘体现未知性、有效性和实用性的特征。数据挖掘技术在应用之前已对大量的数据加以分析整合,通过数据挖掘,可以预测未来会发生的行为。数据挖掘技术在运用中常融入了其他学科,如演绎、归纳等。数据挖掘融合了机器学习和数据库管理。数据挖掘结果可以用于决策分析。在档案领域,数据挖掘可以从大量数字档案中获取到可被利用信息,使档案管理更加高效,可以为用户提供更加优质的档案服务。

2 档案数据挖掘现状

2.1 档案数据管理方式落后

当前许多研究机构、企业以及国家相关部门仍采用较为落后的档案管理方式,不同部门之间独立管理,不同类型档案文件缺乏关联,这使得在档案管理中浪费了大量的人力物力资源。随着信息技术的快速发展,档案管理开始积极吸取各种先进的管理经验,呈现多元化发展态势,这也从侧面反映出档案数据挖掘技术的重要性。

2.2 档案信息化管理系统不够成熟

目前,部分档案管理人员缺乏对现有信息系统可操作性的详细考察,没有结合软件开发的实际情况,制定信息化技术的应用步骤,导致信息化系统的安全性与稳定性难以得到有效维护。一些档案信息化管理系统的建设虽然进行了需求分析,但

未能对档案信息化管理工作的主要原则进行总结,导致动态化管理方法难以在档案管理中实现普及推广。一些档案信息化管理系统在具体建设的过程中,没有结合系统功能特点进行数据模型的创新设计,难以保证信息化技术在档案管理中发挥作用。一些档案管理信息系统的建设、档案管理技术的设计缺乏对系统建设过程的创新调整,导致档案管理的基础性工作和信息技术难以得到高质量的结合,无法在档案管理业务模式创新方面取得积极成效,也使得档案信息化管理系统所具备的应用价值无法实现高水平开发。

3 数据挖掘技术在档案信息管理系统中的实现

3.1 档案资源的分类

档案管理要实现分类,是指依据特定的规则将档案划分不同的种类,将同属性、同一类档案集要,分类是实现整理共组的基础。档案分类可以借助数据挖掘技术中的决策树算法,按照一定的规则划分档案,对档案加以归纳,分区不同类型的档案。档案归类后,档案的检索效率与速度得以提升。数字挖掘技术应用于档案分类要依靠系统数据库的属性分析功能,将资源划分为不同类别。具体方法为:收集特征各异的档案集合,挑选出已分类的训练集。依据训练集应用到的数据挖掘分类方法,依据模型对数据资料加以分类。在档案管理中,数据挖掘分类技术的应用可以帮助用户在多样化的信息中找到想要的信息。数据挖掘要结合信息使用者的需求,针对数字档案资源的数据挖掘,且档案服务要树立协同理念,消除档案信息资源服务方面存在的界限,推动边界融合,构建资源共享体系。在实施中,要重视网络技术的应用,整合档案资源,基于用户需求,推进共建共享。

3.2 数据管理层管理

各种方式与渠道挖掘的数据结构多样、种类繁多,通过数据管理层的处理、集成与存储,可实现以下功能:其一,将结构复杂、难于管理的数据转化为便于处理、结构单一的数据。现代互联网环境下,非结构化数据与半结构化数据应用幅度不断增长,这些资源数据逐渐成为馆藏重要来源;其二,对挖

掘数据采取“清洗”与“去噪”处理，以此保障数据的可靠性与高质量。由于刚采集数据有冗余数据以及噪声数据，对数据进行解析、清洗以及重构，能够有效地提升挖掘数据质量；其三，储存数据。为不同的数据建立独立的数据库并进行存放，缩短访问与查询时间，提升数据获取效率。

3.3 系统测试分析

为判断此次信息检索以及信息挖掘技术应用可靠性以及有效性，此次借助于Hadoop系统基础实现了档案管理系统，涵盖了单个云计算节点、8个的数据库节点以及7个的云存储节点的构建，系统的事务频繁性数据集的搭建利用Centos6.2开源信息库实现，基于acci-dents数据口为测试数据口，利用大数据挖掘方案，把数据划定为一百组，那么总的传达结果即589，档函信息检索以及数据挖掘食品模拟分析综合之前仿真设置来实现，并基于最低制度情况阈值转变，档案信息分析于应用各种策略期间，其加速比对比结果反映此次策略加速较为理想。结合这一测试充分反映出，大数据并行闭频繁项集挖掘应用此次策略期间，可以较好的达成档案数据检索，应用节点彼此的输入输出的传输带宽损耗也进一步降低，由此使得档案数据检索的加速比显著提升，反应较好的加速特性，能够很好的满足大数据背景下的档案数据系统的学习检索以及挖掘等方面的诉求。

3.4 提高档案信息资源的整合水平

在制定档案管理信息化方案过程中，需要加强对档案管理中信息孤岛问题的重视和分析，尤其要对档案信息资源浪费问题进行严格统计，使信息系统可以有效避免重复建设问题，并保证信息匮乏现象可以得到充分规避。要加强对档案资源质量特征的分析考察，并对档案信息管理以及档案服务体系建设的相关需求进行总结，确保不同类型的信息化系统可以与档案信息调取应用需求相符合，有效提升档案服务综合质量。要加强对档案管理中信息化系统实际运行情况的关注，并对财务管理和资源调度管理系统进行整合，结合档案管理团队办公自动化的目标，制定档案信息资源整合的具体策略，使档案数字化管理体系的建设可以获得高质量的信息源，并保证档案信息的来源能够得到相应的优化。在档案信息资源整合的具体实施过程中，一定要加强对档案管理信息化总体方案的关注，尤其要对数字化档案管理优势进行总结，使各种类型的存量信息实现高质量整合，为档案信息的备份和调取提供有力支持。

3.5 实现档案资源整合

针对档案信息资源，要实现其文化价值还要考虑到不同用户的实际需求，要保证“供”与“需”处于平衡状态，要建立立足于市场条件下的平衡机制。档案内容的利用还要结合

民生，民生问题直接对接城市居民。为实现档案资源的共享价值，档案利用要结合城市建设与发展进程的需要，档案供给要转变传统的方式，要尽可能地延伸到社会的多个群体，要尽可能地采用多种方式扩大档案资源的服务范围。数据挖掘要确保档案资源服务与用户的需求加以匹配，要保证精准判断用户的个性化需求。数字档案服务者要借助大数据技术、精准地判断档案资源用户的个性需求，以提供精准性的服务，用户在获取到服务推荐后，会查询、浏览、下载，系统将用户的行为信息加以记录并反馈给算法推荐系统，系统结合反馈信息调整服务策略，实现了服务的优化。

3.6 提高档案数据检索工作质量

在探索信息技术在档案管理工作中的具体应用策略过程中，一定要加强对档案信息构成情况的分析，并结合档案信息查阅需要，对档案信息检索途径进行优化，使档案检索工作目录得到创新调整，进而满足档案信息的多元化查阅需要。档案数据检索工作在开展的过程中，必须加强对信息技术普及规律的重视，并将各类主题词及相应的检索内容加以总结，使档案信息利用需求可以得到优化，有效提升档案检索工作效率。要加强对可视化分析技术的重视和应用，尤其要对图表形式的信息进行重点研究，使数据检索精准度得到提高。档案数据检索工作还需要对特殊信息进行集中处置，尤其要对数据分析模型进行灵活使用，为档案信息利用效率的优化提供精准支持。

结语

综上所述，现阶段的档案数据挖掘还存在着一些问题亟待解决，这对于档案管理工作的可持续发展有着一定的阻碍作用。基于这种状况，很多相关从业者开始探索基于大数据技术的档案数据挖掘，目的是进一步提升档案管理信息化进程，解决现有的档案数据挖掘问题，实现档案管理的长远发展。

参考文献

- [1] 陈小莉. 基于大数据的计算机数据挖掘技术在档案管理系统中的应用[J]. 信息与电脑, 2018(2): 142-145.
- [2] 胡瑛. 大数据处理技术在地理信息档案管理中的应用[J]. 现代测绘, 2016(05)
- [3] 张玉梅. 基于电子信息技术的档案备份与管理[J]. 办公室业务, 2017(06): 168-169.
- [4] 唐洁. 大数据时代下医院档案信息管理对策及开发利用[J]. 西藏医药, 2014, 35(01): 54-55.
- [5] 栾立娟, 卢健, 刘佳. 数据挖掘技术在档案管理系统中的应用[J]. 计算机光盘软件与应用, 2015(01).
- [6] 张文元, 张倩. 大数据技术与档案数据挖掘[J]. 档案管理, 2016(02).