

基于大数据的大庆油田图书馆特色文献文化基因挖掘与分析

李佩

大庆油田文化集团油田图书馆，黑龙江 大庆 163000

DOI:10.61369/SDR.2025030007

摘 要： 大庆油田图书馆作为我国重要的石油工业文献资源中心，其特色馆藏蕴含着丰富的工业文化基因。随着数字技术的发展，运用大数据分析对馆藏文献进行系统挖掘已成为可能。通过构建文献知识图谱，可以清晰呈现石油工业技术发展的历史脉络与知识关联。这种基于数据的文化基因分析方法，不仅能揭示特色文献的内在价值，更能为工业文化遗产保护提供新的研究视角。图书馆应充分利用自身资源优势，在数字人文背景下探索文献资源开发的新路径。

关 键 词： 大数据；大庆油田图书馆；特色文献文化基因

Mining and Analysis of the Cultural Genes of Characteristic Documents in Daqing Oilfield Library Based on Big Data

Li Pei

Oilfield Library of Daqing Oilfield Culture Group, Daqing, Heilongjiang 163000

Abstract： As an important center of petroleum industry literature resources in China, the Daqing Oilfield Library has a characteristic collection that contains rich industrial cultural genes. With the development of digital technology, it has become possible to systematically mine the library's collection of documents by using big data analysis methods. By constructing a literature knowledge graph, the historical context and knowledge connections of the technological development in the petroleum industry can be clearly presented. This data-based cultural gene analysis method can not only reveal the intrinsic value of characteristic literature, but also provide a new research perspective for the protection of industrial cultural heritage. Libraries should make full use of their own resource advantages and explore new paths for the development of literature resources in the context of digital humanities.

Keywords： big data; Daqing oilfield library; characteristic literature and cultural genes

引言

特色文献资源的文化基因研究是图书馆学领域的新兴方向，对专业性图书馆的馆藏建设具有重要意义。大庆油田图书馆保存的石油工业文献具有鲜明的地域特色和行业特征，其内容包含大量反映我国石油工业发展的关键信息。通过大数据技术的应用，可以从文献计量、内容挖掘、知识发现等多个维度，系统分析这些特色资源的学术价值和文化内涵。这种跨学科研究方法既能提升专业图书馆的服务水平，又能为行业历史研究提供可靠的数据支持。

一、大数据与大庆油田图书馆特色文献概述

（一）大数据技术的概念与特点

大数据技术是指处理海量、多样化、快速变化数据的信息处理技术体系。其核心特征体现在四个方面：处理对象具有规模庞大的数据量，远超传统数据库的处理能力；数据类型呈现结构化与非结构化并存的多样性特征，包括文本、图像、音频等多种形式；数据处理过程强调时效性，要求在可接受时间内完成分析任务；技术架构具备水平扩展能力，可通过分布式计算实现性能

提升。在具体实现层面，大数据技术包含数据采集、存储、处理、分析和可视化等完整技术链条。该技术体系已经发展出包括Hadoop、Spark等在内的多种技术框架，并形成了一套完整的方法论体系。^[1]

（二）大庆油田图书馆特色文献的定义

大庆油田图书馆特色文献是指该馆收藏的以石油工业为主题的专业文献资源集合。这些文献在内容上集中反映我国石油工业发展历程，特别是大庆油田开发建设过程中的重要技术成果和管理经验。从载体形式看，包括纸质图书、期刊、技术报告、会议

作者简介：李佩（1986.02-），女，辽宁昌图人，大学本科，中级职称，研究方向：图书馆地方文献。

录、专利文献等多种类型。时间跨度覆盖从油田开发初期至今的完整历史时期。地域特色鲜明，重点收藏与大庆油田直接相关的第一手资料。专业性强，内容涉及地质勘探、钻井工程、油气开采、炼油化工等石油工业全产业链技术领域。这些文献具有不可替代的史料价值和专业参考价值。

二、文化基因挖掘的理论与方法

（一）文化基因的内涵

文化基因是指承载特定文化信息的核心元素，具有可传递性和持久性特征。从内容构成看，包括价值观、行为规范、技术传统等多个维度。在表现形态上，既体现为物质文化载体，也存在于非物质文化遗产形式中。其传播过程遵循特定的演化规律，会随社会环境变化而发展变异。识别标准包括稳定性、影响力和代表性等特征。在石油工业领域，文化基因具体表现为特有的技术传统、管理理念和行业精神等要素。^[2]

（二）适用于特色文献的文化基因挖掘方法

针对特色文献的文化基因挖掘需要采用专门的方法体系。内容分析法通过对文献主题和关键词的系统梳理，识别核心文化要素。共现分析技术可以揭示不同文化要素之间的关联关系。时空分析方法能追踪文化基因的演变轨迹。语义网络构建技术可直观展示文化基因的结构特征。比较研究方法有助于识别特色文献中独特的文化基因。这些方法需要根据文献特点和挖掘目标进行有机组合，形成针对性的分析方案。在具体实施过程中，还需要建立标准化的编码体系和分析框架，确保研究过程的规范性和结果的可比性。

（三）大数据在文化基因挖掘中的优势

大数据技术为文化基因挖掘提供了新的技术支撑。处理能力方面，可以高效处理海量文献资源，突破传统方法的规模限制。分析维度方面，支持多角度、多层次的交叉分析，发现深层的文化关联。技术集成方面，能够结合自然语言处理、机器学习等多种技术手段。可视化呈现方面，可直观展示复杂的文化基因网络。实时分析方面，实现对新产生文献的即时处理。智能化程度方面，通过算法自动识别潜在的文化基因模式。这些技术优势使文化基因挖掘的深度和广度得到显著扩展，为文化研究提供了全新的技术路径。^[3]

三、大庆油田图书馆特色文献文化基因挖掘实践

（一）特色文献数据的收集与预处理

特色文献数据的收集与预处理是文化基因挖掘的基础环节，需要建立系统化的工作流程。在文献收集阶段，重点针对三类资源开展系统采集：馆藏纸质文献的数字化处理、已建数字资源的整合入库以及外部相关文献的补充收集。预处理过程包含四个关键步骤：元数据标准化处理确保文献描述的一致性；文本清洗消除数字化过程中产生的识别错误；内容结构化处理将非结构化文本转化为可分析的格式；知识单元标引对文献内容进行主题标记。技术实现采用OCR识别、自然语言处理等技术手段，构建专用预处理工具链。

质量控制环节设立多重校验机制，包括自动校验规则和人工抽检相结合的方式。通过建立完整的文献数据仓储，为后续分析提供高质量的输入数据。该阶段的规范化程度直接影响最终挖掘结果的可靠性，需要投入足够的技术资源和人力保障。

（二）文化基因的识别与提取

文化基因的识别与提取需要构建科学的分析方法体系。采用多层次分析框架，首先通过词频统计和共现分析识别高频核心概念；其次运用主题模型挖掘文献中的潜在主题分布；然后结合语义网络分析揭示概念间的关联关系。具体操作中，对技术类文献侧重提取专业术语和创新方法，对管理类文献聚焦组织模式和价值理念，对人物传记类文献提炼精神特质和行为范式。提取过程采用计算机辅助与专家研判相结合的模式，先由算法生成候选文化基因集合，再经领域专家评估确认。建立标准化的提取流程文档，详细记录每个文化基因的出处依据和判定理由。为保障结果的可追溯性，对每个提取的文化基因标注来源文献和具体位置。通过这种方法确保提取结果既具有数据支撑，又符合专业知识判断。^[4]

（三）文化基因的分类与编码

文化基因的分类与编码需要建立系统化的知识组织体系。分类框架设计采用多维度标准，包括内容维度（技术、管理、精神等）、形式维度（显性、隐性）和时间维度（传统、现代）。编码系统包含三级结构：大类标识采用字母编码，小类使用数字序列，具体基因实例附加特征标签。分类过程实施双重校验机制，先由算法进行初步归类，再由专家团队审核调整。为每个文化基因建立完整元数据记录，包括定义描述、典型例证、关联基因等属性。编码规则注重可扩展性，预留足够的分类空间适应新增内容。建立分类与编码的质量控制流程，确保同类基因归并合理，异类基因区分明确。最终形成的分类编码体系既反映石油工业特色，又符合通用的知识组织规范，为后续的文化基因分析和应用奠定基础。

（四）文化基因挖掘结果的验证与评估

文化基因挖掘结果的验证与评估需要建立多维度的评价体系。内容效度检验采用专家评议法，邀请石油工业史、文化研究和图书馆学领域的专家组成评审组，对提取的文化基因进行专业审定。结构效度分析运用统计方法检验分类体系的内部一致性。实用价值评估通过试应用方式，观察文化基因在展览展示、教育培训等实际场景中的适用效果。评估指标设置科学性、完整性和实用性三个维度，每个维度下设具体的评价细项。验证过程采用迭代优化模式，根据评估结果对挖掘方法进行针对性改进。建立文化基因知识库的版本管理机制，记录每次修改的内容和依据。通过持续的验证与评估，不断提升文化基因挖掘结果的质量，使其能够准确反映大庆油田特色文献的文化内涵。这一过程需要保持严谨的学术态度，确保最终成果具有足够的权威性和可靠性。^[5]

四、基于大数据的大庆油田图书馆特色文献文化基因分析与应用策略

（一）文化基因的特征分析与规律总结

大庆油田图书馆特色文献文化基因的特征分析与规律总结需

要采用多维分析方法。在时间维度上，可以观察到文化基因具有明显的代际传承特性，早期形成的技术规范和管理理念在后期文献中持续体现。内容维度分析显示，文化基因呈现核心－边缘分布结构，少数关键基因占据主导地位，多数基因形成补充网络。传播方式呈现出从技术文献向普及读物扩散的特征，专业术语逐步演变为通用概念。演变规律方面，技术类基因更新速度较快，精神类基因则保持相对稳定。关联性分析发现不同类别的文化基因之间存在协同演化关系，技术创新往往带动管理理念的更新。稳定性评估表明，体现石油工业本质特征的基因具有更强的持续性。通过对这些特征的深入分析，可以建立文化基因演化的理论模型，为后续的开发利用提供科学依据。分析方法综合运用了统计分析、网络分析和时间序列分析等多种技术手段。

（二）基于文化基因的特色文献资源整合

基于文化基因的特色文献资源整合需要构建系统化的组织框架。整合策略采用“基因－文献－知识”三级关联模式，以文化基因为节点建立文献网络。资源组织按照基因分类体系构建多维导航结构，支持按技术流派、发展阶段等多路径检索。知识关联通过语义技术实现基因之间的智能链接，揭示深层的知识关系。可视化呈现采用时空地图和概念网络等形式，直观展示文化基因的分布与演变。系统架构设计包含数据层、分析层和应用层，确保整合过程的完整性和扩展性。元数据标准专门制定文化基因标引规范，统一描述基因属性和文献关联。用户接口提供个性化定制功能，允许研究者按需组合文化基因视图。这种基于文化基因的资源整合方法，实现了从文献管理到知识组织的转变，为特色文献的深度开发建立了新的范式。

（三）文化基因在图书馆服务创新中的应用

文化基因在图书馆服务创新中的应用体现在多个业务领域。知识服务方面，基于文化基因开发专题知识导航，提供精准的知识推送服务。阅读推广活动围绕核心文化基因设计主题展览和读书会，增强传播效果。教育培训项目根据不同受众需求，开发基于文化基因的系列课程。参考咨询服务利用基因关联网络，提升专题咨询的回答深度和广度。数字人文研究支持学者开展文化基

因的量化分析和可视化展示。文创产品开发提取典型文化基因元素，设计具有石油特色的文化创意品。空间服务通过文化基因主题展示区，营造特色文化氛围。智慧服务应用人工智能技术，实现文化基因的自动识别和智能推荐。这些应用创新不仅拓展了图书馆服务的深度和广度，也提升了特色文献资源的社会价值。实施过程中需要建立跨部门协作机制，确保各项服务的专业性和持续性。^[6]

（四）特色文献文化基因传承与发展的策略

特色文献文化基因传承与发展的策略需要建立系统化的实施体系。资源建设方面，持续补充新产生的特色文献，完善文化基因库的覆盖范围。技术保障方面，应用大数据和人工智能技术提升基因挖掘和分析的效率。人才培养方面，组建跨学科的专业团队，提升文化基因研究能力。合作机制方面，与高校、科研机构建立协同研究网络，开展深度合作。传播渠道方面，构建线上线下结合的立体传播体系，扩大文化基因的影响力。创新研究方面，支持基于文化基因的学术探索和应用开发。制度建设方面，制定文化基因挖掘与应用的工作规范和标准体系。效果评估方面，建立科学的评价指标，持续改进工作方法。这些策略的实施需要长期规划和分步推进，既要保持文化基因的本真性，又要促进其创新发展。通过系统化的策略体系，确保大庆油田特色文献文化基因得到有效传承和合理利用。^[7]

五、结束语

大数据技术为专业图书馆的特色文献研究开辟了新的路径，通过对大庆油田图书馆馆藏资源的系统分析，可以深入挖掘其中的文化基因价值。未来研究应进一步完善数据处理方法，优化分析模型，提升研究结果的准确性和可靠性。同时，要加强跨学科合作，探索文献资源开发与利用的更多可能性。这种基于数据的文化基因分析方法不仅有助于专业图书馆的特色化建设，也为相关行业的创新发展提供了重要的知识支撑。

参考文献

- [1] 张鹏辉. 图书馆服务企业科技创新的实践与思考——以大庆油田图书馆为例[J]. 中国石油企业, 2024, (10): 107-109.
- [2] 周锐. 网络信息化背景下图书馆多元管理模式探究——以大庆油田图书馆为例[J]. 黑河学刊, 2021, (02): 109-111.
- [3] 周锐. 企业图书馆社会化服务的探索与实践——以大庆油田图书馆为例[J]. 大庆社会科学, 2021, (01): 136-138.
- [4] 朱东惠. 企业数字图书馆建设刍议[J]. 大庆社会科学, 2019, (04): 119-120.
- [5] 朱东惠. 基于油田企业文化构设的图书馆阅读推广研究[J]. 大庆社会科学, 2019, (02): 141-142.
- [6] 王迪, 张鹏辉. 图书馆为社会服务职能的延伸——以大庆油田图书馆“匠心·尚学堂”社会公益培训为例[J]. 河南图书馆学刊, 2018, 38(12): 107-108.
- [7] 商亚梅. 大数据环境下图书馆精准服务研究——以大庆油田图书馆为例[J]. 大庆社会科学, 2018, (02): 152-153.