



开源音乐数智图书馆的构建框架与路径研究： 语义化与智能化的集成^{*}

唐振贵 李秀华 杨立志

摘要 站在数字化、语义化和智能化的纵深集成视角,文章旨在探讨通用的音乐数智图书馆(MDIL)构建框架与实现路径。从音乐资源的元数据和内容层面切入,文章以多维度语义组织、AI双向赋能、系统化与模块化、开源优先四大原则为指引,涵盖预处理、知识组织和智能应用三个阶段,并细化功能、数据与技术三条路径,构建描述音乐资源元数据、歌词、乐谱和音频的集成本体模型,结合大模型、检索增强生成和智能体等开发 MDIL 原型系统。以广西原生态民歌数智图书馆的构建为实践案例,实现民歌资源的多维度检索和智能化应用。结果表明该框架是 MDIL 实践层面的系统性框架,结合其开放共享理念,可成为新时期实现特色音乐资源创造性转化和创新性发展的有效途径。

关键词 音乐数智图书馆 语义化 智能化 开源 原生态民歌

分类号 G250.7

DOI 10.16603/j.issn1002-1027.2026.04.008

引用本文格式 唐振贵,李秀华,杨立志. 开源音乐数智图书馆的构建框架与路径研究:语义化与智能化的集成[J]. 大学图书馆学报,2026,44(4):65-76.

1 引言

音乐资源作为一种专门且特色的馆藏资源,在图书馆的资源组织与利用中有其独特的地位,通常涵盖音乐及与音乐相关的各种载体资料^[1]。然而,受限于音乐资源的专业性和技术处理的复杂性,图书馆通常沿用普通文献类型的著录方式对其进行信息组织,缺乏对专业内容的深度加工与有效揭示。进入数智时代,学界与业界积极探索智能技术在音乐资源组织及音乐数字图书馆建设中的应用,相关研究主要集中于数字化、语义化、智能化等较为分散的理论阐述和概念设计,尚缺乏深入的实践研究,尤其是系统性的构建框架与路径研究,且实践成果的开放共享也相对不足。2025年8月21日,国务院发布《关于深入实施“人工智能+”行动的意见》,明确提出支持智能化研发工具和平台的推广应用,探索普惠高效的开源应用模式^[2]。在此背景下,本文尝试从框架设计、技术实现和实践应用三个层面,探

索 AI 赋能的音乐资源的组织和利用,提出一个通用的开源音乐数智图书馆(Music Digital Intelligence Library, MDIL)的框架与构建路径。MDIL 是指利用智能技术深入音乐资源作品元数据和内容层面进行知识组织,实现音乐资源的数字化、数据化和智能化,旨在促进数智环境下音乐资源转化和利用的知识服务平台。本文详细阐述 MDIL 的构建框架与实现路径,并基于开源理念开发 MDIL 原型系统,以广西民歌音乐资源数智化为案例开展应用实践,以期为我国音乐数智图书馆的建设、传统音乐文化的数字化保护与创新传承提供“人工智能+”时代的参考和借鉴。

2 相关研究

总体来看,音乐图书馆的建设正经历从数字化、网络化向语义化、数据化和智能化的升级,其资源组织方式也从元数据描述逐步拓展至内容层面的知识

^{*} 广西哲学社会科学课题研究课题“广西边境民族地区原生态民歌数字化保护与传播路径研究”(编号:23FMZ027)的研究成果之一。
通讯作者:唐振贵,邮箱:blackjack_edu@hotmail.com。



挖掘,实现领域知识的深度揭示与关联,推动馆藏音乐资源由文献服务向知识服务转型。

2.1 音乐资源的数字化

自20世纪90年代以来,国内外音乐图书馆积极推进馆藏资源数字化和应用^[1]。具有代表性的国际项目包括“贝多芬数字档案(Beethoven Digital Archive)”“中世纪音乐数字图像档案(Digital Image Archive of Medieval Music, DIAMM)”,以及印第安纳大学开发的数字音乐图书馆开源系统 Variations^[3](基于Java技术实现馆藏音频和乐谱资源的数字化和网络化)。在国内,上海图书馆的老唱片数字化、天津图书馆的音乐图书馆等特色服务,以及多所音乐院校图书馆的音乐资源数字化实践,均体现了这一趋势,研究主题涵盖音乐资源数字化、音乐数据库建设和评价等方面,对音乐资源的描述主要集中在元数据层面^[4-11]。

2.2 音乐资源的语义化

随着语义网技术的兴起,音乐资源的组织方式逐渐向语义化方向演进。国外典型案例包括,法国国家图书馆的 Overture 探索平台,实现古典音乐资源知识图谱和关联数据的语义检索和可视化导航^[12];德国柏林国立图书馆的巴赫音乐图书馆,提供了巴赫音乐的数字化和智能化数据^[13]。英国伦敦城市大学、大英图书馆等多家机构合作开发的数字音乐实验室^[14],构建了面向大规模音乐收藏的研究方法和软件基础设施^[15]。音乐资源的语义化主要涉及元数据语义化和内容语义化两个方面。在元数据语义化方面,IFLA 图书馆参考模型(Library Reference Model, LRM)^[16]与美国音乐图书馆协会的《RDA 元数据指南》^[17]相结合,能够全面完整地描述音乐作品及其内容表达,实现音乐资源元数据的语义描述和关联数据发布。Music Ontology^[18]、Music Meta Ontology^[19]和 DOREMUS^[12]等音乐描述本体从异构资源集成的角度描述音乐资源,支持不同来源音乐资源元数据的互操作。在内容语义化方面,MusicNote^[20]、MusicOWL^[21]本体用于描述乐谱符号和结构;Segment 本体^[22]描述音乐结构;Chords 本体^[23]、Functional Harmon 本体^[24]描述合弦,WASABI 本体^[25]可用于歌词标注,AFO(Audio Feature Ontology)本体、AFV(Audio Feature Vocabulary)词汇^[26]、Audio Commons 本体^[27]等用于描述音乐音频特征。以上本体丰富了音乐资源

内容的语义化表达,推动音乐资源内容层面在语义检索、知识图谱构建和关联数据发布方面的研究和应用^[12,25,28-29]。

相较而言,国内学者在音乐资源语义化方面的研究主要集中在元数据的语义化处理^[30-31],虽有部分研究涉及音乐作品的内容语义层面^[32-33],但主要停留于设计和构想阶段,尚未深入开展实践应用。学者们认为应当充分借鉴国际上先进经验和技术方法,深入音乐资源的内容层面,推动我国的音乐资源的数字化和语义化升级^[1,6,32]。

2.3 音乐资源的智能化

“数智”赋能是图情档发展的新增长点^[34],随着图书馆逐渐向数字化、信息化、智能化转型进阶,数智图书馆(或称智慧图书馆)已成为图书馆当下和未来的发展方向和建设目标^[35-36]。尤其在图书馆资源组织的数智转型与知识生产方面,“数智”强调以人工智能技术拓展资源组织范围,深化资源融合能力、推动数智技术应用^[37]。音乐资源的数智化升级是图书馆数智化愿景的重要体现,智能技术在音乐资源内容识别、音乐信息检索和创造性的智能应用等方面提供了强有力的支撑。(1)在内容识别方面,乐谱作为音乐作品的音乐信息载体,是内容识别的重要对象。Audiveris^[38]、Moonlight^[39]、Orchestra^[40]、Mozart^[41]等光学乐谱识别工具可从图像中识别乐谱符号并输出为结构化数据;MT3^[42]等音轨转录开源模型能够将音频文件转录为 MIDI 文件,该格式文件可进一步转换为乐谱文件;此外,Open-SheetMusicDisplay^[43]、Verovio^[44]、MuseScore^[45]等开源工具为乐谱的呈现和编辑提供跨平台的解决方案。(2)在音乐信息检索方面,大语言模型凭借其对自然语言和音乐知识的理解能力,提升传统音乐信息检索的性能,并扩展出自然语言检索、检索增强生成、跨模态检索等智能化检索方式^[46-48]。(3)在创造性的智能应用方面,音乐资源可作为大语言模型的训练数据或提示词样例,用于提升大模型的音乐理解能力,有助于构建具有音乐分析和音乐创作能力的音乐大模型^[49-51];音乐知识库可深度嵌入音乐智能体的认知决策框架,成为其领域知识来源,参与智能体的自主决策和工具调用等环节,支撑智能体完成歌曲创作、音乐分析等高阶认知任务^[52-53]。

综上所述,我国音乐资源数字化和音乐数字图书馆的研究已取得一定成效,但从“人工智能+”与



“开源”国家政策视角进行审视,图书馆普遍缺少足够的人力、财力和技术处理专业性极强的音乐资源,导致在音乐资源的语义化,尤其是内容层面的语义化与智能化方面的研究深度不够。尽管国外研究已积累大量音乐数智图书馆建设所需的开放本体模型和开源软件,智能技术的发展在一定程度上降低了音乐知识处理的专业门槛,但图书馆在融合语义化、智能化和开源工具方面仍面临诸多挑战:(1)音乐资源的内容涉及多学科知识,尚缺乏跨领域建模的集成本体;(2)智能技术多聚焦于音乐信息处理的单一环节,尚缺乏面向数智图书馆所需的通用框架与路径,更缺少可供参考的面向图书馆服务需求的智能应用案例;(3)开源软件系统往往采用不同的技术栈、异构的运行环境,加大了各环节集成的难度。

3 音乐数智图书馆的构建框架和实现路径

3.1 MDIL 的构建框架与设计原则

针对当前音乐数字图书馆(MDIL)在上述方面存在的难点与不足,本文从设计和实现层面展开研究,

探索音乐数智图书馆的构建框架及其实现路径。MDIL 构建框架旨在通过音乐资源的数字化、语义化和智能化,为不同应用场景的用户提供知识服务。具体而言,面向普通读者提供音乐资源的元数据检索服务;面向专业读者提供内容层面的领域知识服务;面向智能环境中的用户,利用知识图谱、大模型和智能体技术整合元数据和领域知识,提供智能交互的知识服务。MDIL 的构建框架如图 1 所示,涵盖四大原则、三个阶段和三条路径,具体设计原则如下。

3.1.1 多维度语义组织原则

实现 MDIL 目标的前提是对音频资源作深度、细致的知识组织,需采用多维度语义组织策略。从元数据和内容层面,应结合图书情报学、音乐学、语言学等领域知识,对文本、图像、音频和视频等形态的音乐资源,以及歌词、乐谱和音频等不同内容表达形式的音乐资源进行语义描述,将知识组织的粒度细化至音乐资源内容中的字符和音符,揭示其音乐学、语言学的多维属性特征,构建音乐资源的知识图谱和关联数据,为 MDIL 在数智环境中的应用奠定坚实基础。

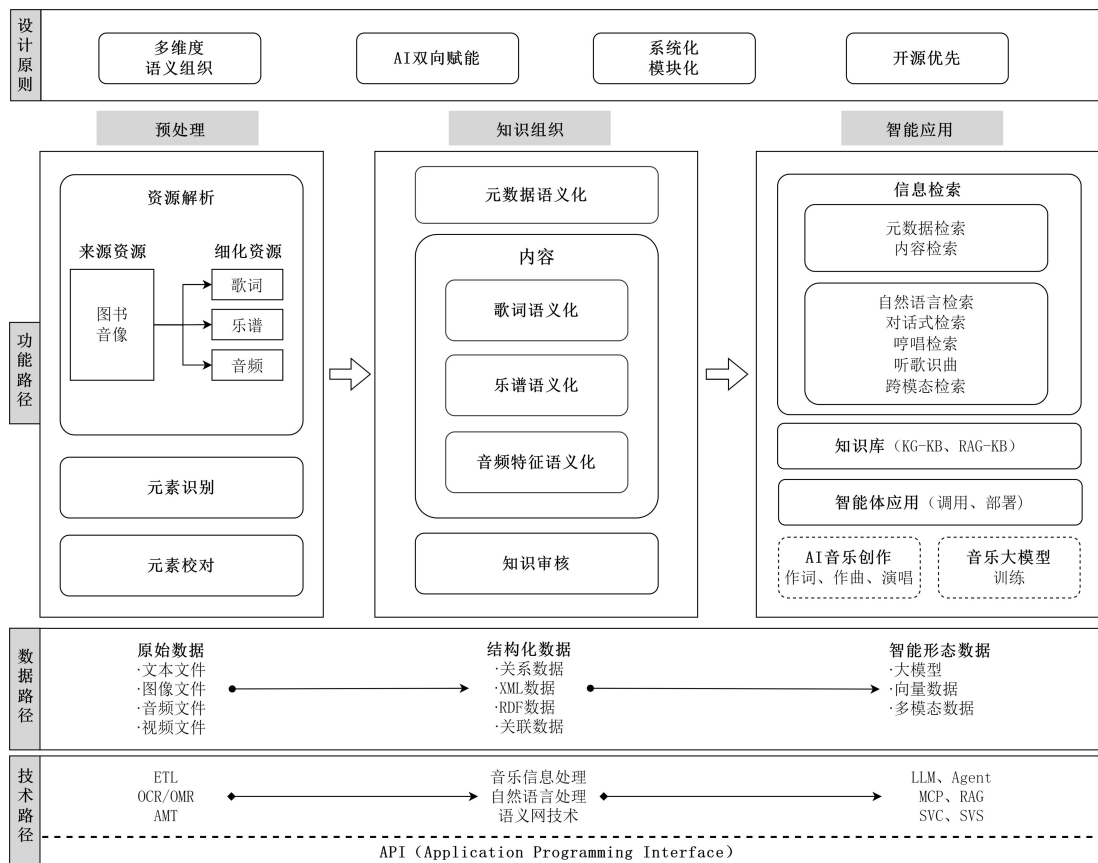


图 1 音乐数智图书馆的构建框架



3.1.2 AI 双向赋能原则

MDIL 强调的智能性体现在“AI 增强”和“增强 AI”两个方面。“AI 增强”是指将 AI 技术融入 MDIL 构建的各个环节,通过“AI+”补充非专业人员的学科领域知识,提升其音乐处理的能力和效率,学科专家则关注知识的检验和审核,形成“学科馆员+AI+学科专家”的协同知识加工形式。“AI 增强”为传统数字图书馆提供新的功能增长点,扩展出语义检索、自然语言检索和跨模态检索等智能检索功能。“增强 AI”指 MDIL 构建的知识服务能反哺 AI 音乐的研究和应用,为 AI 音乐大模型的训练提供优质语料,为智能体、深度研究(DeepResearch)等 AI 应用提供丰富、专业和特色的知识来源。“AI 增强”和“增强 AI”相辅相成、相互促进,共同实现 MDIL 的 AI 双向赋能。

3.1.3 系统化和模块化设计原则

MDIL 采用系统化和模块化设计思路,以指导其软件的开发和实现。系统化设计将音乐资源的采集、加工、组织和智能化应用视为一个相互关联的完整流程。模块化设计按“高内聚、低耦合”的原则,通过标准接口进行通信,并通过模块组合构建系统。系统化和模块化使得 MDIL 通过灵活配置满足不同项目规模和应用场景的需求,同时降低集成开源系统的难度。

3.1.4 开源优先原则

在 MDIL 构建框架中优先采用成熟的开源项目,一是有效降低 MDIL 软件开发和运维成本,使预算有限的机构也能受益;二是提高 MDIL 的可扩展性,以适应不同应用场景的需求,应对技术快速迭代环境中的功能扩展。例如,图书馆可利用乐谱识别、音乐信息处理和资源语义化等免费开源项目降低成本投入,按图书馆业务流程进行二次开发,扩展检验、审核和回溯等功能,提高音乐数据的质量与关联外部知识的能力。

3.2 MDIL 构建框架的实现路径

遵照 MDIL 的设计原则,可将 MDIL 实现路径从逻辑上划分为预处理、知识组织和智能应用三个阶段,从类型上区分为功能路径、数据路径和技术路径,形成“3×3”的推进矩阵。各阶段的功能模块采用适配器和装饰器等模式设计 API 架构^[54],集成内部功能模块并对接外部开源系统。本节以功能路径为主线,结合技术路径和数据路径,阐述 MDIL 框架的实现路径。

3.2.1 预处理阶段

预处理阶段的主要任务是解析音乐资源和识别音乐符号元素,为知识组织阶段准备不同粒度的组织单元。其中,(1)资源解析功能模块负责从记录音乐作品的来源文献中抽取音乐作品个体,再从音乐作品中解析出歌词、乐谱和音频等细化资源,例如从包含音乐作品的图书文献(含随书资源)中抽取音乐作品的歌词、乐谱和音频文件;从磁带、光盘等介质中转录或提取音乐作品的音频文件。(2)元素识别模块聚焦于细化资源的内容信息,识别音乐内容表达的符号元素,如歌词(文字符号)、乐谱符号和音频信号。从技术路径而言,细化资源的预处理采用 ETL(Extraction Transformation Loading)技术实现,歌词文本和乐谱符号通过基于图像识别的 OCR(Optical Character Recognition)和 OMR(Optical Music Recognition)技术实现,针对音频对象,通过 AMT(Automatic Music Transcription)技术将音乐内容(如旋律、和弦、节奏等)转换为 MIDI 数据,再转换生成乐谱符号。目前,上述识别技术均无法达到 100% 的准确性,仍需进行人工校对,确保抽取与识别出的内容与原始文献一致。

3.2.2 知识组织阶段

知识组织阶段是 MDIL 实现路径的核心阶段。该阶段对预处理阶段输出的来源资源、细化资源及其内容元素作语义化组织,从多层次、多学科角度描述音乐作品,构建语义关联数据,为应用阶段提供数据支撑和知识来源。其功能路径可细分为两个层面。第一,在元数据语义化层面,MDIL 采用 LRM 模型描述音乐资源的元数据,通过使用和扩展 WE-MI(Work、Expression、Manifestation、Item)实体的子类 and 关系,描述来源资源与细化资源基本结构和逻辑关系,并采用音乐图书馆协会(MLA)扩展的 RDA 词汇进一步描述资源在音乐方面的特征。第二,在内容语义化层面,MDIL 以歌词、乐谱和音频三种细化资源作为知识组织的对象。(1)歌词语义化主要描述歌词文本的语言和音乐信息。语言信息包括词性、词汇、语义等,以及多语言歌词涉及的语言翻译信息;音乐信息包括歌曲结构(主歌、副歌、桥段、前奏、间奏、尾奏)、韵律结构等。(2)乐谱语义化是在乐谱数字化的基础上,利用乐谱相关的本体模型描述音符、音高、时长、旋律、力度、调性等乐谱信息,促进机器对音乐符号的理解,实现不同记谱法和编码



标准之间的数据转换,使乐谱资源更易于检索和发现。(3)音频特征语义化是从音频的物理特征与感知特征等角度描述音乐作品的声音文件,如峰值、能量、强度等时域或频域特征,以及节奏、音高和调式等音乐特征。音频中蕴含的音乐特征描述能支持音乐信息学中数据的语义关联与集成,有助于音乐分析和音乐检索。

在技术路径和数据路径上,音乐资源知识组织的核心任务是音乐数据的结构化处理,目标是使音乐资源更易于检索、发现和利用。MDIL 对音乐内容的结构化分为两个步骤:第一步,使用自然语言处理、音乐信息处理工具对音乐作品的文字、图像、音频资源进行识别和解析,获得第一轮结构化数据,如 XML、JSON 等结构化文档以及关系数据库的关系数据结构,这类数据在一定程度上依赖于特定的软件或解析器读取,尚未实现跨领域的知识集成;第二步,采用语义网技术,通过不同领域本体模型描述音乐资源,将第一步生成的结构化数据转换为 RDF 数据和关联数据,进一步增强数据的语义理解和逻辑推理能力,通过与外部知识库关联形成更大的知识网络,有助于结构化数据在智能环境中的灵活利用。

知识组织阶段还设有知识审核模块。与预处理阶段的元素校对不同,知识审核关注的是知识的有效性和准确性。元素校对主要检查所识别元素与原始文献的一致性,为保持数据来源的可回溯性,对原始文献的印刷错误、记谱过程中存在知识性错误等问题,并不直接勘误或修订。此外,受限于自动化技术的可靠性,数据标注和转换也可能存在一定偏差,以上因素共同影响数据质量。知识审核通过学科专家验证、高性能 AI 评估以及本体模型的语义约束与推理规则验证,系统化识别并修正知识错误,从而确保知识准确性并提升数据质量。

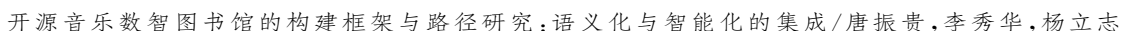
3.2.3 智能应用阶段

智能应用阶段是集成音乐资源的语义数据,通过智能技术呈现各类功能应用。MDIL 提供基础、增强和扩展三类功能模块,通过不同模块的集成灵活适配多种应用场景,例如侧重于获取和访问原始资源的检索系统、支持知识图谱可视化呈现的知识库、对接大模型的智能体等典型应用。

基础功能模块以信息检索为核心,为用户提供多种检索方式,以获取和访问各类音乐资源,包括元数据检索、内容检索和融入智能交互技术的检索。

元数据检索基于结构化的描述性信息检索资源,而内容检索的对象是音乐作品的内容资源(歌词文本、乐谱和音频),两者均为信息检索系统最基本的检索形式。融入智能交互技术的检索包括,基于自然语言文本交互的自然语言检索,支持历史记录、多轮对话的对话式检索,基于音频输入的哼唱检索(Query by Humming, QBH)、基于歌曲片段进行的听歌识曲(录音检索),以及融合文本、乐谱、音频等多模态的跨模态检索。其中,基于音频输入的检索和跨模态检索尤其契合音乐资源的多媒体特性。

增强类功能模块为音乐资源的多种知识形态应用提供支撑,包括面向专业读者服务的知识库和面向智能环境的智能体应用。在知识库方面,MDIL 知识库按技术实现方式分为基于知识图谱的知识库(Knowledge Graph-based Knowledge Base, KG-KB)和基于检索增强生成的知识库(Retrieval-Augmented Generation-based Knowledge Base, RAG-KB)。KG-KB 通过图数据库存储音乐资源的 RDF 三元组数据,支持用户使用 SPARQL、Cypher 等专用查询语言进行语义检索,实现知识及其关联关系的精准获取。KG-KB 不仅具备自动推理机制以增强知识发现能力,还能促进异构信息系统间的互操作性,其可视化分析功能为音乐知识的探索与呈现提供了直观的交互方式。RAG-KB 是面向大模型智能交互场景应用的知识库,其核心是通过 RAG 技术将音乐作品的元数据、歌词、乐谱和音频特征等数据按照不同策略进行知识重组,利用嵌入模型构建高维度的向量特征,形成可供大模型直接使用的外挂知识库。相较于 KG-KB, RAG-KB 更适于实现人机对话,能够利用大模型的语言理解能力来获取和分析音乐资源,生成符合人类阅读习惯的文本。在智能体方面,MDIL 利用智能体的自主规划和自动执行能力增强知识服务,基于智能体的应用通过大语言模型分析用户请求,制定相应策略自动从信息检索模块、知识库模块中获取资源,并决定是否调用外部资源或执行外部工具来优化结果,将各类资源整理、归纳和总结后返回给用户,进而根据用户反馈动态调整策略。智能体应用包含调用和部署两种模式,调用模式是指 MDIL 调用外部智能体来扩充自身功能,例如调用个人助手、个性化推荐等智能体构建音乐图书馆的 AI 馆员;部署模式是将 MDIL 的语义描述、数据转换、信息检索等功能封装成



扩展类功能模块以 MDIL 数据作为高质量、特色化的语料数据,支持 AI 音乐创作和音乐大模型训练。具体而言,借助歌声转换(Singing Voice Conversion, SVC)和歌声合成(Singing Voice Synthesis, SVS)等技术,对歌词、乐谱和音频数据再创作,生成新的音乐资源。该类功能通常对算力资源有较高要求,属于数智图书馆可选配的特色增值模块。

MDIL 构建框架作为实现音乐数智图书馆的通用性基础指南,适用于面向各类特色音乐资源的数智图书馆构建。为此,本节从 MDIL 的本体设计、技术实现和实践案例三方面阐述 MDIL 的构建过程,并以具有地方民族特色的广西原生态民歌数智图书馆作为实践案例。

MDIL 本体模型作为知识组织阶段的核心框

架,主要目标是在知识层面集成各类异构音乐资源,并与已有外部领域本体进行关联。在图书馆信息资源整合中,常用的本体设计方法包括单一本体法、多本体法和混合方法^[55],本文采用混合方法设计 MDIL 本体模型(图 2)。首先,基于 IFLA 的 LRM-er 本体^[56]进行扩展,设计专门用于描述音乐资源及其歌词、乐谱和音频细化资源的元数据模块;其次,围绕各细化资源的具体内容设计内容模块,并通过映射方式与外部领域本体建立关联。在元数据模块中,音乐资源内容表达类(MusicResExpression)继承自 LRMer 的内容表达类(lrmer:E3_Expression),通过其乐谱内容表达(MusicalNotationExpression)、歌词内容表达(LyricExpression)和音频特征表达(AudioExpression)子类,描述细化资源在内容表达层面的元数据信息,并通过具体化属性(lrmer:R3_isEmbodiedIn)关联至预处理阶段抽取的文档、图片或音频文件等载体文件。在内容表达层面,采用互逆属性“hasFragment”和“isFragment-Of”,建立来源资源与细化资源之间的关联。

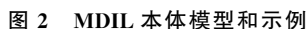




图2以《壮族民歌100首》^[57]中的作品《壮乡三月风光美》为例,展示MDIL本体描述音乐资源的RDF三元组的代码片段,其中,片段1是音乐作品元数据描述的示例。在本体设计中,每个内容表达类均配有专注描述内容本身的内容类,从而清晰地区分了内容表达的元数据描述和内容描述。具体而言,乐谱内容表达类侧重于元数据描述,而乐谱内容类(MusicalNotationContent)则描述由某种记谱方法记录音乐作品的符号元素,乐谱内容类按记谱方法可进一步细分为五线谱内容(StaffContent)、简谱内容(JianpuContent),以及便于字符表示和机器解析的ABC谱内容(ABCNotationContent)。其中五线谱内容直接映射到外部本体MusicOWL的mso:Score类,以MusicXML标准实现五线谱语义化描述,图2“片段2”是五线谱三元组的示例。歌词内容类(LyricContent)描述文本结构及其语言和音乐方面的语义信息。其中,TextUnit类精细描述歌词文本的位置和片段,通过字(Character)、词(Word)、句(Sentence)、段(Paragraph)等子类及其关系构建文本结构,并通过“hasLinguisticInformation”和“hasMusicalInformation”属性,分别关联至文本的语言信息类(LinguisticInformation)和音乐信息类(MusicalInformation)。图2的示例“片段4”描述歌词文本“三(哪的)月(类)山(哪)歌(哪)声声甜”相关语言信息(如词性、命名实体),以及民歌中的“衬词”的音乐信息。音频内容类由音频特征对象类(AudioFeatureObject)体现,通过“hasAudioFeature”属性关联具体的特征,而对特征的描述则重用外部领域本体实现,如图2示例“片段3”,使用AFO和AFV本体定义的音乐音频特征类“afv:MFCC”描述音频在某个时间间隔中的梅尔频率倒谱系数^①。

4.2 技术实现

MDIL的原型系统^[58]基于开源技术构建,采用多语言协同开发策略。前端基于JavaScript与HTML5实现响应式交互界面,集成知识图谱可视化组件以支持异构数据的直观展示;后端采用Python语言与FastAPI框架构建轻量级RESTful服务,确保高效的数据处理与接口通信。数据存储层结合关系

型数据库PostgreSQL与图数据库Neo4j,分别实现结构化数据的持久化与知识图谱的语义关联存储。智能体模块依托LangChain框架开发,通过集成大语言模型实现自然语言理解与任务自动化推理,采用Streamlit开发智能体前端应用。系统通过模块化设计实现技术组件的解耦,并利用容器化技术提升部署灵活性,最终形成支持多模态数据融合与智能分析的原型系统。

4.3 应用案例

广西原生态民歌作为我国非物质文化遗产的重要组成部分,承载着丰富的地域文化、民族情感与历史记忆,兼具极高的艺术价值与学术研究价值。然而,当前与其相关的文献资源建设与管理仍面临突出挑战:一方面,相关资源分布零散、载体多元、结构异质,图书馆、博物馆和档案馆等文化机构对其组织多停留于书目层面的联机检索,尚未能实现内容层面的深度关联与语义组织;另一方面,由于原生态民歌在音乐形态上的独特性以及语言上的多样性,大量特色资源尚未得到有效揭示与充分利用,导致学术研究、文化传承与创新应用之间存在显著鸿沟。在此背景下,如何依托数字化与智能化技术,对广西原生态民歌相关的各类文献与音像资源的内容进行系统化、结构化的知识组织,构建细粒度、可互操作的语义关联和智能服务体系,已成为推动该领域学术发展、实现文化遗产可持续保护与活态传承的关键课题。鉴于此,本文以广西壮族自治区图书馆地方文献库、网络公开资源及田野调查采集的原生态民歌音乐资源作为主要数据来源,以MDIL构建框架为指导,通过MDIL本体模型将分布零散、形态多样的音乐资源进行细粒度的知识组织,并基于大语言模型、RAG和智能体等技术实现智能应用,形成数字化、智能化于一体的知识服务平台(以下简称平台)。当前阶段收集的来源资源包含音乐作品1826首,数字化文件6325个(包含来源资源和细化资源),元数据9632条,语义化阶段的RDF数据约85万个三元组,智能化阶段产生的合成数据约15GB,图3展示了单首音乐作品导入平台的具体步骤。

^① 梅尔频率倒谱系数(Mel-Frequency Cepstral Coefficients, MFCC)是一类常用音频特征,用于表征音色/谱包络,广泛应用于语音与音频分析。在AFV本体中,afv:MFCC是对MFCC特征的本体化建模,用作特征类型的语义标识,以统一描述并关联某段音频(或某一帧/片段)上计算得到的MFCC系数及其相关元数据。

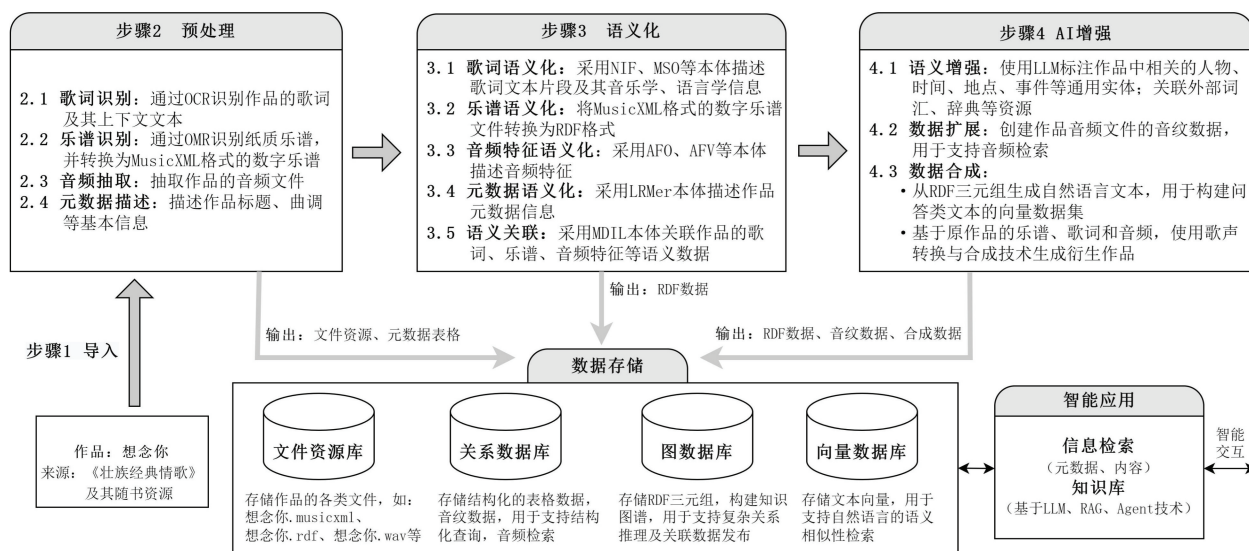


图3 作品导入平台的具体步骤

平台的特点在于实现广西原生态民歌资源的元数据检索、内容检索和智能应用。相较于传统以元数据为中心的检索模式,本平台的核心优势在于深度融合数字化与智能化技术的内容检索能力,支持专业用户在音乐分析、文化研究与创作应用等场景下,实现精准、高效的资源发现与语义级内容关联。本节从前端用户交互的角度,介绍该平台在内容检索方面提供的五项数字化与智能化功能,包括AI检索、多语言检索、乐谱检索、录音检索、语义检索。图4展示平台检索功能的整体页面及AI检索交互功能,各检索功能的界面及技术实现详见本研究的开源代码仓库^[58]。

(1) AI检索功能

AI检索功能采用基于自然语言理解的对话交互形式,以“AI馆员”智能体为载体,实现直观的人机交互。“AI馆员”能够精准识别并解析用户意图,自动调用民歌数据资源与平台功能接口,动态响应不同用户的检索需求。如图4所示,用户可通过上传录音发起检索请求,“AI馆员”将调用平台录音检索接口匹配音乐作品,并进一步推荐基于生成式技术合成的衍生作品,从而拓展音乐资源的呈现形式与传播途径。

(2) 多语言检索功能

多语言检索功能旨在突破原生态民歌唱词的語言多样性壁垒。原生态民歌的歌词通常采用多样化的语言文字记录形式(如壮文、古壮字、汉语等),具

有鲜明的民族地域特色,平台通过歌词语义化,将不同语言的歌词文本进行对齐,并关联至方言词汇库与专业辞典资源,实现歌词的跨语言检索和语义检索。当用户以歌词中的壮文词汇“daeng ngoenz(太阳)”作为检索条件时,平台可基于语义对齐与跨语言映射机制,返回包含汉语、壮语等多语言文本形式的关联结果。

(3) 乐谱检索功能

乐谱检索功能以数字乐谱资源的深度组织与高效利用为核心,支持用户通过上传乐谱图像或直接录入乐谱符号的方式发起检索。该功能依据用户提交的乐谱片段,在资源库内进行内容级特征匹配与定位,最终向用户返回与之相对应的完整乐谱或匹配度较高的乐谱局部片段。

(4) 录音检索功能

录音检索功能突破了传统图书馆联机公共目录查询系统(OPAC)以文本为中心的检索模式,借助音频相似度分析技术,为音乐类资源提供了更契合其媒介本质的检索途径。用户可通过上传或实时录制音频片段,对平台中的音乐作品进行基于音频内容相似度的识别与检索,为普通用户实现“听歌识曲”的功能。对于专业研究需求,平台不仅能够返回匹配的音乐作品,还可进一步定位至作品中相对应的具体片段,并提供波形图等可视化分析视图,辅助用户开展进一步的音乐结构与特征分析。



图4 广西原生态民歌数智图书馆的AI检索功能界面

(5) 语义检索功能

语义检索功能基于语义网技术构建,通过对歌词、乐谱与音频资源的语义数据进行统一建模与关联,实现对音乐内容的概念化检索。针对具备专业知识的研究者,平台支持直接使用 SPARQL 或 Cypher 等查询语言进行精准、灵活的语义检索与知识探索。为降低使用门槛,平台通过集成大语言模型,将用户自然语言表达的查询意图转换为规范的查询语句。用户通过构建 SPARQL 语句执行音符检索,其返回结果不仅支持三元组数据导出,还可通过交互式可视化界面进行导航,辅助用户探索相关资源的深层语义关联。

在广西原生态民歌数智图书馆的设计与实现过程中,MDIL 框架为平台设计提供系统化、模块化、流程化的设计思路。通过 MDIL 框架梳理清晰的功能路径、数据路径与技术路径,有助于精准把握平台的功能需求、数据需求和技术选型。同时,MDIL 框架提供的开源工具,是实现平台低成本开发的关键,这些工具通过提供标准化组件和模块,有效提升开发效率,降低技术门槛与项目实施成本。广西原生

态民歌数智图书馆经过为期六个月的运行,已完成核心功能模块的系统集成,系统运行状态稳定可靠。经课题组语言学、音乐学专家的协同验证,平台基本满足原生态民歌资源语义化和智能化的功能需求。

5 结语

本研究集语义化、智能化与开源理念,提出 MDIL 构建框架及实现路径,通过细粒度的资源整合、多维度的语义化知识组织和智能化的应用服务,为图书馆的知识服务升级提供了创新性解决方案,同时为音乐文化遗产的数字化保存与创新性转化构建了系统性框架,能够较好满足传统音乐在数字人文语境下的保存与再生产需求。在实践层面,开源的 MDIL 原型系统为音乐知识服务的智能化转型提供低成本、可复制的技术范式,并通过模块化设计与标准化接口,为学界与业界开展跨机构应用实践与功能扩展提供了开放共享的技术平台。

本研究尚存在如下不足。其一,在传统民族音乐的数字化实践中,MDIL 本体仅初步实现简谱与五线谱标准(MusicXML)及本体(MusicOWL)的部



分映射规则需求,完整的简谱本体模型和解析器仍有待进一步研究完善。鉴于《音乐曲谱出版》系列国家标准即将颁布实施,MDIL 下一阶段将据此开发完整的简谱本体模型及其转换框架,更全面地支持民族音乐资源的数据语义化,为中国传统音乐融入全球数字人文生态提供有效途径。其二,现阶段的 MDIL 原型系统主要实现了核心功能模块的基础架构,初步验证了系统设计的完整性与可行性。然而,受研究经费和版权限制,本文尚未开展大规模性能评估,未来研究计划与相关机构合作,在更大范围的应用场景中开展实践研究。

参考文献

- 王建欣,杨媛. 数字人文背景下音乐图书馆音乐文献资源数字化工作研究[J]. 图书馆工作与研究,2021(11):69-75.
- 国务院. 国务院关于深入实施“人工智能+”行动的意见 [EB/OL]. [2025-10-01]. https://www.gov.cn/gongbao/2025/issue_12266/202509/content_7039598.html.
- Indiana University Digital Library Program. Variations3 [EB/OL]. [2025-10-01]. <https://dlib.indiana.edu/projects/vari-ations3/index.html>.
- 温带宝. 特色资源建设与专业服务整合研究——以音乐图书馆为例[J]. 图书馆学研究,2012(6):70-73.
- 张黎静. 音乐数据库系统的设计与实现[D]. 武汉:华中科技大学,2013:11.
- 孙宇. 数字人文背景下音乐图书馆音乐文献资源数字化建设研究[J]. 图书馆工作与研究,2023(3):70-75.
- 王建欣,杨媛. 数字人文背景下音乐图书馆音乐资源数字化工作研究[J]. 图书馆工作与研究,2021(11):69-75.
- 马英珺. 中国音乐大典数据库的设计开发与应用[J]. 中国音乐,2021(6):10-14.
- 冯晨晨. “中国音乐大典数据库”建设意义探析[J]. 数字人文,2023(2):83-100.
- 曾粤亮,吕晓龙,朱明珠. 以资源为中心:数字记忆平台建设评价指标体系构建[J]. 图书馆学研究,2024(5):53-67.
- 刘宇统,张成香. “库中库”全共享式民歌数据库建设基础和核心构想[J]. 淮阴师范学院学报(哲学社会科学版),2023,45(1):92-98.
- Achichi M, Lisena P, Todorov K, et al. DOREMUS: a graph of linked musical works[C]//International Semantic Web Conference. Cham: Springer International Publishing, 2018: 3-19.
- Bach Digital. Welcome to Bach Digital [EB/OL]. [2025-10-01]. <https://www.bach-digital.de/content/index.xed>.
- Digital Music Lab. About [EB/OL]. [2025-10-01]. <https://dml.city.ac.uk/>.
- Cottrell S. Big music data, musicology, and the study of recorded music: three case studies[J]. The Musical Quarterly, 2018, 101(2-3): 216-243.
- IFLA FRBR Review Group. IFLA Library Reference Model: a conceptual model for bibliographic information[EB/OL]. [2025-10-01]. <https://repository.ifla.org/items/214c74cb-c075-4428-a138-39f8d06c55aa>.
- MLA Content Standards Subcommittee. Music Library Association RDA metadata guidance (Version 0.93) [EB/OL]. [2025-10-01]. <https://wp.blog.musiclibraryassoc.org/mla-best-practices/>.
- Motools. Music ontology specification[EB/OL]. [2025-10-01]. <https://github.com/motools/musicontology>.
- De Berardinis J, Carriero V A, Mero no-Peñuela A, et al. The music meta ontology: a flexible semantic model for the interoperability of music metadata[PP/OL]. V1. arXiv(2023-11-07) [2025-10-01]. <https://doi.org/10.48550/arXiv.2311.03942>.
- Cherfi S S, Guillotel C, Hamdi F, et al. Ontology-based annotation of music scores[C]//Proceedings of the 9th Knowledge Capture Conference. 2017: 1-4.
- Jones J, De Siqueira Braga D, Tertuliano K, et al. Musicowl: the music score ontology[C]//Proceedings of the International Conference on Web Intelligence. 2017: 1222-1229.
- Fields B, Page K, De Roure D, et al. The segment ontology: bridging music-generic and domain-specific[C]//2011 IEEE International Conference on Multimedia and Expo. IEEE, 2011: 1-6.
- Mauch M, Dixon S. Simultaneous estimation of chords and musical context from audio[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2009, 18(6): 1280-1289.
- Kantarelis S, Dervakos E, Kotsani N, et al. Functional harmony ontology: musical harmony analysis with description logics[J]. Journal of Web Semantics, 2023, 75: 100754.
- Fell M, Cabrio E, Tikat M, et al. The WASABI song corpus and knowledge graph for music lyrics analysis[J]. Language Resources and Evaluation, 2023, 57(1): 89-119.
- Allik A, Fazekas G, Sandler M. Ontological representation of audio features [C]//International Semantic Web Conference. Cham: Springer International Publishing, 2016: 3-11.
- Ceriani M, Viola F, Rudan S, et al. Semantic integration of audio content providers through the audio commons ontology[J]. Journal of Web Semantics, 2023, 77: 100787.
- Semantic-audio. AudioGraph [DB/OL]. [2025-10-01]. <https://semantic-audio.github.io/>.
- Ratta M, Daga E. Knowledge graph construction from musicXML: an empirical investigation with SPARQL anything [EB/OL]. [2025-10-01]. https://oro.open.ac.uk/85326/1/Music_Knowledge_Graphs_Paper%20%283%29.pdf.
- 曹建军,李俊萱,王雨荷. 中国传统音乐知识库的“元数据本体”构建研究[J]. 数字人文研究,2022,2(4):22-37.



- 31 陈雅玲,何琳. 基于开放关联数据的音乐资源整合研究[J]. 图书馆杂志,2018,37(9):86-95.
 - 32 杨媛. 数字人文视域下图书馆音乐资源平台建设研究[J]. 图书馆工作与研究,2022,(7):89-96.
 - 33 赵季若. 数字人文视角下音乐类档案资源知识组织研究[J]. 山西档案,2024,(2):119-121.
 - 34 孙建军,李阳,裴雷.“数智”赋能时代图情档变革之思考[J]. 图书情报知识,2020,(3):22-27.
 - 35 苑刚,曲明娟. 数智技术赋能智慧图书馆建设研究[M]. 北京:新华出版社,2024:1.
 - 36 白如江,张新雨,牛湘荷. 数智化环境下高校图书馆未来学习中心建设现状与发展框架研究[J]. 图书情报工作,2025,69(7):90-101.
 - 37 樊振佳,夏翠娟. 知识诸宝是求:数智时代的图书馆资源建设与知识组织[J]. 中国图书馆学报,2025,51(2):22-34.
 - 38 Audiveris. Audiveris open-source optical music recognition[EB/OL]. [2025-10-01]. <https://github.com/Audiveris/audiveris>.
 - 39 TensorFlow. Moonlight optical music recognition (OMR) [EB/OL]. [2025-10-01]. <https://github.com/tensorflow/moonlight>.
 - 40 AdelRizq. Orchestra [EB/OL]. [2025-10-01]. <https://github.com/AdelRizq/Orchestra>.
 - 41 Aashrafh. Mozart :convert sheet music to a machine-readable version[EB/OL]. [2025-10-01]. <https://github.com/aashrafh/Mozart>.
 - 42 Magenta. MT3: multi-task multitrack music transcription[EB/OL]. [2025-10-01]. <https://github.com/magenta/mt3>.
 - 43 OpenSheetMusicDisplay. OpenSheetMusicDisplay (OSMD): a musicXML renderer for the browser[EB/OL]. [2025-10-01]. <https://github.com/opensheetmusicdisplay/opensheetmusicdisplay>.
 - 44 Verovio. A music notation engraving library[EB/OL]. [2025-10-01]. <https://www.verovio.org/index.xhtml>.
 - 45 MuseScore. MuseScore: music notation and composition software[EB/OL]. [2025-10-01]. <https://github.com/musescore/MuseScore>.
 - 46 Qin Y, Xie H, Ding S, et al. Score images as a modality: enhancing symbolic music understanding through large-scale multimodal pre-training[J]. Sensors (Basel, Switzerland), 2024, 24(15): 5017.
 - 47 Wu S, Guo Z, Yuan R, et al. Clamp 3: universal music information retrieval across unaligned modalities and unseen languages[PP/OL]. V1. arXiv(2025-02-14)[2025-10-01]. <https://doi.org/10.48550/arXiv.2502.10362>.
 - 48 Le D V T, Bigo L, Herremans D, et al. Natural language processing methods for symbolic music generation and information retrieval: a survey[J]. ACM Computing Surveys, 2025, 57(7): 1-40.
 - 49 Ma Y, oland A, Ragni A, et al. Foundation models for music: a survey[PP/OL]. V1. arXiv(2024-08-26)[2025-10-01]. <https://doi.org/10.48550/arXiv.2408.14340>.
 - 50 Lin Y, Dai Z, Kong Q. MusicScore: a dataset for music score modeling and generation [PP/OL]. V1. arXiv(2024-06-17)[2025-10-01]. <https://doi.org/10.48550/arXiv.2406.11462>.
 - 51 Wang Y, Wu S, Hu J, et al. Notagen: advancing musicality in symbolic music generation with large language model training paradigms[PP/OL]. V1. arXiv(2025-02-25)[2025-10-01]. <https://doi.org/10.48550/arXiv.2502.18008>.
 - 52 Yu D, Song K, Lu P, et al. Musicagent: an AI agent for music understanding and generation with large language models[PP/OL]. V1. arXiv(2023-10-18)[2025-10-01]. <https://doi.org/10.48550/arXiv.2310.11954>.
 - 53 Tatar K, Pasquier P. Musical agents: a typology and state of the art towards musical metacreation[J]. Journal of New Music Research, 2019, 48(1): 56-105.
 - 54 Gough J, Bryant D, Auburn M. Mastering API architecture: design, operate, and evolve API-based systems[M]. O'Reilly Media, Inc, 2021.
 - 55 欧石燕. 语义网与数字图书馆[M]. 南京:南京大学出版社, 2017:113-114.
 - 56 IFLA. LRMer (IFLA Library Reference Model Entity Relationship) [EB/OL]. [2025-10-01]. <https://www.iflastandards.info/lrm/lrmer.html>.
 - 57 范西姆. 壮族民歌 100 首[M]. 南宁:广西民族出版社, 2009:3-7.
 - 58 Blackjackedu. Music digital intelligence library [EB/OL]. [2025-10-01]. <https://github.com/blackjackedu/MDIL>.
- 作者贡献说明:
唐振贵:设计研究思路与框架、论文撰写、系统开发
李秀华:论文撰写与修改
杨立志:论文案例设计
- 作者单位:唐振贵、李秀华,广西财经学院新闻与文化传播学院,广西南宁,530003
杨立志,浙江音乐学院图书馆,浙江杭州,310024
- 收稿日期:2025 年 10 月 11 日
修回日期:2025 年 12 月 8 日
- (责任编辑:王菲)



Research on the Construction Framework and Pathways of an Open-Source Music Digital Intelligence Library: Integrating Semantic and Intelligent Approaches

TANG Zhengui LI Xiuhua YANG Lizhi

Abstract: Music resources constitute a specialized and distinctive category of library holdings. Due to their professional nature and the complexity of associated technical processing, the information organization of music resources has long depended on conventional bibliographic description methods, which fail to provide in-depth content processing or effective disclosure of specialized materials. In the era of digital intelligence, addressing these limitations and enabling intelligent, service-driven applications requires libraries to develop reusable solutions across knowledge modelling, processing frameworks, and system integration. The study approaches music resources from both metadata and content levels, adhering to four guiding principles: multi-dimensional semantic organization, bidirectional AI empowerment, systematization and modularization, and open-source prioritization. It encompasses three sequential stages—pre-processing, knowledge organization, and intelligent application—and further refines three interrelated pathways of functionality, data and technology. An Music Digital Intelligence Library (MDIL) integrated ontology model is constructed to describe music resource metadata, lyrics, musical scores, and audio. Based on this ontology model, a knowledge graph is built, and by integrating large language models, retrieval-augmented generation, AI agents, and open-source technologies, an MDIL prototype system is developed. At the practical level, the construction of the “Guangxi Original Ecological Folk Song Digital Intelligence Library” serves as a case study. Leveraging the MDIL framework and prototype system, multi-dimensional retrieval and intelligent application of folk song resources are realized. On one hand, the system supports metadata integration and organization of heterogeneous resources; on the other, it provides content retrieval functions including AI retrieval, multilingual retrieval, musical score retrieval, audio recording retrieval, and semantic retrieval. Empirical results demonstrate that the MDIL enhances the efficiency of organizing, disclosing, discovering, and utilizing folk song resources, lowers the technical threshold and project implementation costs of the practical pathway, and forms a systematic solution for upgrading library knowledge services.

Keywords: Music Digital Intelligence Library; Semanticization; Intelligentization; Open-Source; Original Ecological Folk Song