

基于 AIGC 的图书馆智能管理助手开发与应用

——以 AIGC 在图书馆知识服务中的应用为例

张冰, 胡婷婷, 玉泉

(广西壮族自治区科学技术情报研究所, 南宁 530022)

摘要: 探索 AIGC(人工智能生成内容)结合 RAG(检索增强生成)和 Stable Diffusion 模型、AI Agent(AI 智能体)等核心技术开展知识挖掘的方法、工具、技术框架与应用实践,开发与应用于 AIGC 的图书馆智能管理助手,助力产业发展与知识传播。重点研究基于 AIGC 的图书馆智能管理助手的开发过程、模型原理及如何利用 AIGC 技术拓展图书馆知识服务,并进一步融合产业发展需求,展望未来服务拓展功能。通过开发与实践,构建集高效、智能、个性化于一体的图书馆智能管理助手,探索 AIGC 技术在图书馆知识服务中的创新应用,以期为大模型在图情领域的专业化、场景化应用提供参考。

关键词: AIGC(人工智能生成内容); 图书馆; 智能管理助手; 知识服务

中图分类号: G250.76 **文献标志码:** A **文章编号:** 1671-1807(2025)11-0077-07

随着人工智能技术的飞速发展,特别是人工智能生成内容(artificial intelligence generated content, AIGC)技术的广泛应用,图书馆知识服务正迎来前所未有的变革机遇。本文开发了“基于 AIGC 的图书馆智能管理助手”的 AI Agent(AI 智能体)应用,利用大模型、AIGC、RAG(retrieval-augmented generation, 检索增强生成)、工作流等核心技术,为用户提供高效、便捷、个性化的图书馆知识服务,同时服务于广西新能源和六堡茶等特色产业,助力产业发展与知识传播。

1 需求分析与规划

1.1 需求分析

在图书馆知识服务领域,当前面临如下痛点和需求。

(1)信息过载与检索效率低。随着图书馆资源的不断增加,读者在面对海量信息时往往感到无从下手,传统的检索方式效率低下,难以满足快速获取精准信息的需求。

(2)个性化服务缺失。不同读者有不同的兴趣和需求,但图书馆往往无法提供个性化的服务,导致读者的阅读体验不佳,难以吸引和留住长期用户。

(3)知识更新滞后。图书馆的知识更新往往滞后于学术和产业发展,无法及时反映最新的研究成果和技术动态,影响了知识服务的时效性和实用性。

(4)跨领域知识整合困难。在新能源、广西六堡茶等领域,读者往往需要跨领域的知识整合,但图书馆现有的服务往往难以实现这一点,导致读者花费大量时间和精力去自行整合知识。

(5)互动体验不足。传统的图书馆服务往往缺乏交互性,读者难以获得即时的反馈和帮助,影响了服务的质量和效果。

1.2 实践目标

基于以上痛点,确定了图书馆在知识服务方面的主要需求,包括高效的信息检索和推荐、个性化服务、实时知识更新、跨领域知识整合以及增强互动体验。为满足这些需求,开发了“基于 AIGC 的图书馆智能管理助手”的 AI Agent 应用,利用大模型、AIGC、RAG、工作流等先进技术,为用户提供高效、便捷、个性化的图书馆知识服务。

实践目标是整合图书馆的海量资源,并结合新能源领域和广西六堡茶领域的专业知识,实现以下功能。

收稿日期: 2024-11-26

基金项目: 2024 年广西壮族自治区科学技术情报研究所基本科研业务费专项(基研专项 2024J-02)

作者简介: 张冰(1989—),女,河南许昌人,硕士,馆员,研究方向为产业情报;胡婷婷(1982—),女,广西南宁人,硕士,副研究馆员,研究方向为科技情报及产业发展;玉泉(1985—),男,广西南宁人,硕士,助理研究员,研究方向为科技查新与产业发展。

(1)智能检索与推荐。利用 RAG 技术和大模型的语义理解能力,实现精准的信息检索,并提供个性化的图书推荐服务。

(2)AIGC 驱动的内容生成。运用 AIGC 技术自动生成图书馆相关的新闻、报告、指南等内容,并通过文生图技术将复杂知识点以图像形式展现。

(3)实时知识更新与整合。实时抓取学术和产业界的最新动态,确保图书馆知识库的时效性和完整性,并提供一站式多领域知识获取。

(4)互动咨询与答疑。借助大模型的自然语言处理能力,实现与用户的智能对话,提供即时的咨询和答疑服务,并利用 AIGC 技术动态生成针对性的回答。

(5)专业领域知识服务。针对新能源和广西六堡茶等特定领域,通过微调技术优化模型,提供精准和深入的专业知识服务,并生成领域相关的深度报告等。

(6)高效的工作流管理。通过集成工作流技术,实现图书馆内部业务流程的自动化和智能化,提高工作效率,并利用 AIGC 技术优化文档生成、报告编写等环节。

1.3 实践规划

基于需求分析,制定了详细的项目规划,包括技术选型、功能模块设计、数据资源准备等。

(1)技术选型。选择基于 AIGC 技术的深度学习框架作为核心,结合大数据处理技术和自然语言处理技术,构建图书馆智能管理助手。

(2)功能模块设计。设计个性化图书推荐、智能化信息检索、跨模态内容生成、智能营销方案设计、智能化决策辅助等功能模块,满足图书馆在知识服务方面的多元化需求。

(3)数据资源准备。从图书馆的各类资源及产业情报等渠道广泛收集数据,并进行数据清洗、标注等预处理工作,为模型训练和功能开发提供高质量的数据支持。

(4)开发与测试计划。制定详细的开发与测试计划,确保项目按时按质完成,并不断优化和完善助手的功能和性能。

(5)上线与迭代规划。规划助手的上线时间和迭代周期,建立用户反馈机制,根据用户反馈不断迭代和优化助手的功能和性能。

2 核心技术及原理

2.1 AIGC 技术

AIGC 是一种利用人工智能技术自动生成文

本、图像、音频等内容的技术。其原理基于对抗生成网络(generative adversarial network, GAN)对比语言-图片预训练(contrastive language-image pre-training, CLIP)等人工智能技术,通过对大量已有数据的学习寻找规律,并以适当的泛化能力生成内容的生产方式^[1]。Gartner^[2]将 AIGC 定义为一种颠覆性的智能技术,它能够在没有人类经验与思维偏见的前提下创造新的结果。麦肯锡(Mckinsey & Company)认为 AIGC 通过一种接近人类行为与人类进行交互式协作,它可能将彻底改变我们创作内容的方式^[3]。

在图书馆智能管理助手中,利用 AIGC 技术实现实时知识更新与整合功能。通过实时抓取并处理学术和产业界的最新动态,AIGC 技术能够自动生成与图书馆服务领域相关的内容,如新闻摘要、研究报告等,确保图书馆知识库的时效性和完整性。

2.2 RAG 模型

RAG 模型是将存储于外部知识源,如向量数据库中的非参数化知识,结合参数化模型,以可解释、模块化的方式捕获知识^[4]。其原理主要基于向量检索技术和生成模型的结合,首先通过向量检索技术从知识库中快速检索到与输入相关的信息,然后将检索到的信息作为生成模型的输入,生成符合用户需求的文本内容。RAG 通过从外部知识源动态检索相关信息以辅助大模型作答,尤其是在知识密集型任务中,能够显著提高回应的准确性和相关性,减少幻觉问题^[5]。目前,大模型在信息检索^[6]、知识抽取^[7]、知识发现^[8]、知识管理与服务^[9]等方面均展现出了巨大潜力。

在图书馆智能管理助手中,集成了 RAG 模型来实现智能检索与推荐功能。通过向量数据库与大模型的结合,RAG 模型能够快速检索到与用户输入相关的信息,并结合用户的历史行为和偏好生成个性化的推荐列表,提高检索的准确性和效率。

2.3 AI Agent 技术原理

AI Agent 指一种人工实体,它们能够使用传感器感知周围环境,做出决策,然后使用执行器采取行动^[10]。它融合了人工智能、机器学习、自然语言处理等多种技术,能够理解用户指令、分析上下文信息,并据此采取适当的行动。AI Agent 通过持续学习和优化,不断提升自身的智能水平和服务质量。AI Agent 是对科技行业的“Shock Wave”,Agent即将彻底改变人们使用计算机的方式^[11]。

图书馆智能管理助手本质是构建一个 AI agent

的智能体,融入 AIGC, RAG, Stable Diffusion 等技术,根据用户需求,基于大模型自主执行思考、决策并执行相应任务。它作为用户的智能代理,能够深入理解用户的需求和偏好,自动执行各种任务。例如, AI Agent 可以基于用户的查询指令,自动调用 RAG 模型进行智能检索,快速定位并返回相关资源;同时,它还能根据用户的请求,调用 Stable Diffusion 模型生成与图书馆资源相关的图像,如书籍封面、插图等,从而丰富图书馆的数字资源,提升用户体验。 AI Agent 的智能化执行和思考能力,使得图书馆智能管理助手能够高效、准确地响应用户需求,提供个性化的服务。

2.4 Stable Diffusion 模型

Stable Diffusion 是一种基于深度学习的图像生成模型,特别擅长根据文本描述生成高质量、多样化的图像。作为一种稳定的图像生成技术, Stable Diffusion 在系统中被用于将文字描述转化为具体的视觉图像^[12],同时保持图像的多样性和高质量。

在图书馆智能管理助手中,充分利用 Stable Diffusion 模型的图像生成能力。当用户需要为图书馆资源生成图像时,如书籍封面、插图等, AI Agent 会自动调用 Stable Diffusion 模型,并根据用户提供的文本描述生成相应的图像。这些生成的图像不仅丰富了图书馆的数字资源,还为用户提供了更加直观、生动的阅读体验。通过 Stable Diffusion 模型的应用,图书馆智能管理助手进一步提升了其服务的多样性和创新性。

3 开发过程

3.1 数据收集与预处理

数据是 AI 模型训练的基础,对于开发基于 AIGC 的图书馆智能管理助手至关重要。从图书馆的各类资源及产业情报中广泛收集数据,并通过一系列预处理步骤确保数据的质量和准确性。

3.1.1 数据收集

1) 图书馆资源数据收集

(1) 馆藏图书数据。通过图书馆管理系统获取馆藏图书的元数据,包括书名、作者、出版日期、ISBN、摘要、分类标签等。

(2) 电子资源数据。收集图书馆订阅的电子期刊、电子图书、数据库等资源,提取其元数据及内容摘要。

(3) 用户行为数据。收集用户借阅记录、搜索历史、浏览记录等行为数据,以了解用户偏好和

需求。

2) 产业情报数据收集

(1) 行业报告。从相关研究机构、咨询公司获取行业研究报告,提取关键数据和洞察。

(2) 专利数据。收集专利数据库中的专利信息,包括专利标题、摘要、发明人、申请日期等。

(3) 新闻与资讯。从新闻网站、专业媒体收集与图书馆服务领域相关的新闻和资讯,提取有用信息。

3.1.2 数据清洗

(1) 去重。通过比对数据集中的记录,去除重复的数据条目,确保数据的唯一性。

(2) 缺失值处理。检查数据中的缺失值,根据具体情况采用填充、删除或插补^[13]等方法进行处理。

(3) 异常值检测与处理。利用统计学方法检测数据中的异常值,并通过领域知识或人工审核进行修正或删除。

(4) 格式统一。将数据转换为统一的格式,如将日期格式统一为 YYYY-MM-DD,确保数据的一致性。

3.1.3 数据标注

(1) 分类标注。对图书、文章等资源进行分类标注,如学科分类、主题分类等,以便进行精准推荐。

(2) 情感分析标注。对评论、评价等文本数据进行情感分析^[14]标注,区分正面、负面和中性情感,以支持情感计算相关功能。

(3) 实体识别标注。对新闻、专利等文本数据进行实体识别标注,如识别出人名、机构名、技术名词等,为知识图谱构建提供支持。

(4) 关系抽取标注^[15]。标注文本中实体之间的关系,如作者与作品的关系、公司与专利的关系等,以支持知识推理和整合。

3.2 数据存储与知识库构建

数据存储与知识库构建在图书馆智能管理助手中起着至关重要的作用,以下是针对向量数据知识库构建和大模型微调语料数据准备的详细说明及其作用。

3.2.1 向量数据知识库构建

向量数据知识库构建是图书馆智能管理助手实现高效、智能检索与推荐功能的基础,在图书馆智能管理助手中发挥着至关重要的作用。通过向量化处理,将图书馆的各类文本资源(如图书、期刊、报告等)转换为高维向量表示,构建高效的知

库和准备高质量的语料数据,使得系统能够快速理解并处理这些资源,为用户提供精准的信息检索和个性化知识服务,具体说明如下:①向量化处理。采用自然语言处理(NLP)技术、Embedding 模型,将文本数据转换为高维向量表示。这些向量不仅保留了原始文本的语义信息,还使得系统能够通过计算向量之间的相似度来评估文本内容的相关性。②知识库设计。根据图书馆资源的特点和需求,设计合理的知识库结构,包括向量存储格式、索引机制、关系图谱等。通过构建高效的索引机制,如倒排索引、向量空间模型等,实现快速检索和推理,提高系统的响应速度和准确性。③数据存储。将清洗、标注和向量化处理后的数据存储到向量数据库中 Milvus 中。这些数据库针对向量数据的存储和检索进行了优化,能够提供高性能和低延迟的查询服务,为 RAG(检索增强生成)模型等提供丰富的数据支持。

3.2.2 大模型微调语料数据准备

大模型微调语料数据准备是提升图书馆智能管理助手个性化服务和专业领域知识服务能力的关键。通过对语料数据的筛选、增强和存储,可以优化大模型的训练过程,使其更好地适应图书馆知识服务的需求,提供更高质量的推荐和生成内容。具体说明如下:①语料筛选。从清洗和标注后的数据中选择与图书馆知识服务紧密相关的语料,确保语料的质量和相关性。这些语料应涵盖图书馆的各类资源,如图书、期刊、新闻等,以及用户的查询记录、行为数据等。②数据增强。通过数据增强技术,如同义词替换、句式变换等,增加语料的多样性和丰富度。数据增强有助于提升大模型的泛化能力,使其能够处理更多样化的输入,并生成更高质量的内容。③数据存储。将筛选和增强后的语料数据存储为适合大模型训练的 JSONL 格式。这些格式不仅便于大模型的读取和处理,还能提高训练效率,确保微调过程的顺利进行。

3.3 大模型微调

在模型微调与优化阶段,具体采用 llama-factory 框架,并针对特定任务和数据集对预训练模型进行了细致的微调,以进一步提升模型的性能和适应性。以下是具体化的步骤和配置细节。

3.3.1 加载预训练模型

利用 llama-factory 框架加载了开源的 9 B 数据集的大模型。该模型具备丰富的语言知识和强大的推理能力,为后续的微调提供了坚实的基础。

3.3.2 配置微调任务

在 llama-factory 中,根据具体需求配置了微调任务,详细如下。

(1)任务类型。文本分类与生成。

(2)输入输出格式。输入为文本序列,输出为分类标签或生成文本。

(3)数据集。采用特定领域(新能源、广西六堡茶)的标注数据集,包含约 5 000 条样本数据。

3.3.3 微调训练

在配置好微调任务和数据后,利用 llama-factory 启动了微调训练过程,具体配置如下。

(1)学习率。初始学习率设置为 0.000 1,并采用余弦退火(Cosine Annealing)策略进行调整,以在训练后期逐渐降低学习率,避免过拟合。

(2)批次大小。设置为 32,以平衡训练效率和内存占用。

(3)优化器。采用 Adam 优化器,其参数 β_1 和 β_2 分别设置为 0.9 和 0.999。

(4)训练轮数。共进行 10 轮训练,每轮训练后都会在验证集上进行评估。

3.3.4 监控与评估

在微调训练过程中,密切监控了以下性能指标。

(1)准确率。用于评估分类任务的性能。

(2)BLEU 分数^[16]。用于评估生成任务的性能,确保生成的文本与参考文本具有较高的相似度。

(3)损失函数值。采用交叉熵损失函数,实时监控其变化以判断模型是否收敛。同时,每训练完一轮后,都会在验证集上进行评估,确保模型在训练过程中不断向更好的方向优化。

(4)模型保存与部署。当微调训练达到满意的效果后(如验证集上的准确率超过 90%,BLEU 分数超过 30),利用 llama-factory 框架将微调后的模型保存下来,并进行了必要的部署工作。具体步骤包括:①模型保存。将微调后的模型权重保存为 .pth 文件。②模型集成。将保存的模型集成到人工智能图书馆管理助手中,确保能够无缝调用。③性能测试。在实际应用环境中对部署后的模型进行性能测试,确保其能够满足实时性和准确性的要求。

通过采用 llama-factory 框架进行上述具体化的微调训练,不仅显著提高了模型的性能和适应性,还确保了微调过程的可控性和高效性。这在有限的时间和资源内实现高质量的模型微调提供了有力支持。

3.4 AI Agent 的构建

在图书馆智能管理助手中, AI Agent^[17]是实现智能化服务的关键组件。AI Agent 通过集成 AIGC、RAG、Stable Diffusion 等先进技术,能够自主执行思考、决策并响应用户任务,为用户提供个性化、高效的知识服务。

3.4.1 AI Agent 架构设计

AI Agent 的架构设计包括感知层、理解层、决策层和执行层 4 个部分。感知层负责接收用户的输入,如文本、语音等;理解层利用自然语言处理技术解析用户意图,并提取关键信息;决策层根据用户意图和当前环境状态,制定合适的响应策略;执行层则负责具体执行响应任务,如检索、推荐、生成内容等。

3.4.2 AI Agent 技术实现

在技术实现上,采用先进的深度学习框架和算法来构建 AI Agent。具体实现步骤如下。

(1)意图识别与槽位填充。设计并实现意图识别模块和槽位填充模块,用于准确识别用户意图并提取关键信息。这两个模块均采用深度学习算法进行训练和优化。

(2)决策策略制定。根据用户意图和当前环境状态,制定合适的决策策略。采用强化学习算法来训练 AI Agent 的决策能力,使其能够在不同场景下做出最优的决策。

(3)任务执行与反馈。执行层负责具体执行响应任务,并收集用户的反馈。通过不断学习和优化, AI Agent 能够逐渐提升任务执行效率和用户满意度。

3.4.3 AI Agent 与知识库、大模型的交互

AI Agent 与知识库、大模型的交互^[18]是实现智能化服务的关键环节。通过设计合理的接口和交互机制,实现了 AI Agent 与知识库、大模型的高效交互。具体交互方式如下。

(1)与知识库的交互。AI Agent 通过向量检索技术从知识库中快速检索到与用户输入相关的信息。构建了高效的向量数据库^[19],并支持实时更新,以确保检索的准确性和时效性。

(2)与大模型的交互。AI Agent 将检索到的信息以及用户输入传递给大模型,利用大模型的强大生成能力生成符合用户需求的文本内容。同时,大模型还可以为 AI Agent 提供额外的背景知识和推理能力支持。

3.4.4 AI Agent 与 Stable Diffusion 交互

为了实现 Stable Diffusion 的接入,进行以下

工作。

(1)环境搭建与模型部署。搭建适合 Stable Diffusion 运行的环境,并部署预训练的 Stable Diffusion 模型。

(2)接口开发与集成。开发 Stable Diffusion 的接口,并将其集成到 AI Agent 中。这样, AI Agent 就可以根据需要调用 Stable Diffusion 模型进行图像生成。

(3)Agent 调度使用。在 AI Agent 的决策层中,增加对 Stable Diffusion 任务的调度逻辑。当需要生成图像时, AI Agent 会自动调用 Stable Diffusion 接口,并传递必要的参数。

3.5 功能开发与测试

(1)前端界面开发。设计并实现用户友好的前端界面,包括聊天交互,确保用户能够便捷地使用助手的各种功能。

(2)后端服务开发。开发后端服务,包括数据处理、模型推理、结果返回等,确保前端界面能够与后端模型进行高效的数据交互。

(3)系统集成与测试。将各个功能模块进行集成,并进行全面的系统测试,包括功能测试、性能测试、安全测试等,确保系统的稳定性和可靠性。
①意图识别测试。设计了多种用户意图场景,测试 AI Agent 的意图识别准确率和响应速度。通过不断优化模型算法和增加训练数据,提高了意图识别的性能。
②任务执行测试。模拟了多种用户任务,如检索、推荐、问答等,测试 AI Agent 的任务执行效率和准确性。针对发现的问题,及时进行了调整和优化。
③用户反馈收集与改进。在实际使用过程中,收集用户的反馈意见,并针对 AI Agent 的不足之处进行改进。

3.6 上线与迭代

AI Agent 作为图书馆智能管理助手的核心组件,其性能和服务质量将直接影响用户的使用体验。因此,在上线后持续关注用户反馈,并根据反馈进行迭代和优化。同时,利用持续学习技术,不断更新 AI Agent 的知识库和模型参数,以适应图书馆知识服务领域的快速发展和用户需求的不断变化。通过持续学习和优化,旨在打造一个更加智能、高效、贴心的图书馆管理助手,为用户提供更加优质的服务体验。

4 利用 AIGC 技术拓展图书馆知识服务

4.1 个性化图书推荐

基于 AIGC 技术的图书推荐系统,不仅能够根

据读者的阅读历史和兴趣偏好,自动生成个性化的图书推荐列表,还能通过深度学习模型不断优化推荐算法,提高推荐的准确性和相关性。这种智能化的推荐系统能够有效提升读者的阅读体验,帮助读者更快找到感兴趣的书籍,同时促进图书馆资源的有效利用。此外,结合用户反馈和行为数据,系统可以持续学习和进化,实现更加精准的个性化服务。

4.2 智能化信息检索

利用 AIGC 技术,图书馆可以实现对馆藏资源的智能化检索。读者只需使用自然语言提出查询需求,系统便能自动解析并匹配相关资源,显著提高检索效率和用户体验,为读者提供更加便捷、全面的信息服务。同时,智能化检索系统还能根据读者的查询历史和学习习惯,自动调整检索策略,优化查询结果的相关性排序。

4.3 跨模态内容生成

AIGC 技术不仅限于文本内容的生成,还能支持图像等跨模态内容的创作。在图书馆知识服务中,这意味着助手可以生成与书籍内容相关的插图、活动海报等,丰富知识服务的形式和内容。帮助读者更直观、生动地理解知识内容。这种跨模态内容生成^[20]能力将极大地提升图书馆知识服务的吸引力和互动性。

4.4 融合产业发展需求案例

基于 AIGC 的图书馆智能管理助手以 Web 应用的方式进行使用,使用方法如下。

(1)在浏览器输入 <http://122.9.155.236:8088/>。

(2)点击图标则会出现聊天界面,通过文字聊天的方式与图书馆智能管理助手进行交互,帮你解决你提出的问题。

(3)基于 AIGC 的图书馆智能管理助手功能点总结:①分析读者兴趣,推荐相关图书,提供书名、图书简介、对读者的帮助及图书位置;②提供新能源和广西六堡茶领域的知识专业咨询;③用图书馆资源和产业情报,制定营销方案。

结合新能源汽车领域的发展趋势和读者需求,图书馆可以提供针对性的专业知识服务。利用 AIGC 技术,图书馆可以自动收集、整理和分析新能源领域的最新技术动态、政策解读、市场趋势等信息,生成易于理解的报告,为研究人员、企业家和普通读者提供有价值的参考。此外,图书馆还可以与新能源企业合作,共同开发定制化的知识服务产品,满足特定行业的需求。例如,图书馆可以定期发布新能源领域的专题报告,涵盖太阳能、风能、电

动汽车等热门子领域的技术进展和市场动态。同时,通过 AIGC 技术生成的相关图表和图像,可以直观地展示各种新能源技术的优势和应用场景,帮助读者更好地理解 and 把握新能源领域的发展趋势。

针对广西六堡茶这一特色产业,图书馆可以利用 AIGC 技术提取相关知识信息,并提供全面的知识服务。这包括但不限于六堡茶的历史文化、种植技术、加工工艺、品鉴方法、市场趋势等方面的内容。通过自动生成相关图片等多媒体内容,图书馆可以生动地展示六堡茶的特色和魅力,吸引更多读者关注和了解。同时,图书馆还可以与茶农、茶企合作,共同推广六堡茶品牌,促进地方经济发展。图书馆智能助手构建一个关于广西六堡茶的专题知识库,包含六堡茶的起源、发展历程、文化内涵等方面的详细介绍。通过 AIGC 技术,图书馆还可以生成与六堡茶相关的各类多媒体内容,为读者提供全方位、多角度的知识服务。

5 总结与展望

目前,基于 AIGC 的图书馆智能管理助手虽然已经具备了可交换性和学习能力,能够通过 RAG 技术提升知识水平,但在自主成长和视频生成方面仍存在短板。因此,后续规划主要集中在提其自主成长与进化能力,以及拓展其多媒体生成功能特别是视频生成技术。通过这些改进和规划,期望基于 AIGC 的图书馆智能管理助手能够为用户提供更加全面、深入、高效和个性化的知识服务,同时,也将持续关注图书馆知识服务领域的发展趋势,结合广西新能源、六堡茶等特色产业的需求,不断优化和拓展智能管理助手的功能和服务。

参考文献

- [1] 张春春, 孙瑞英. 如何走出 AIGC 的“科林格里奇困境”: 全流程动态数据合规治理[J]. 图书情报知识, 2024, 41(2): 39-49, 66.
- [2] Gartner. What is artificial intelligence[EB/OL]. (2023-06-28) [2024-10-20]. <https://www.gartner.com/en/topics/artificial-intelligence>.
- [3] McKinsey&Company. What is generative AI[EB/OL]. (2023-06-28) [2024-10-20]. <https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai#/>.
- [4] 赵浜, 曹树金. 生成式 AI 大模型结合知识库与 AI Agent 开展知识挖掘的探析[J/OL]. 图书情报知识, 1-14[2024-11-20]. <https://link.cnki.net/urlid/42.1085.G2.20241103.1003.002>.
- [5] GAO Y F, XIONG Y, GAO X Y, et al. Retrieval-aug-

- mented generation for large language models: a survey [EB/OL]. (2024-01-01) [2024-10-20]. <https://arxiv.org/abs/2312.10997v5>.
- [6] 赵鑫, 窦志成, 文继荣. 大语言模型时代下的信息检索研究发展趋势[J]. 中国科学基金, 2023, 37(5): 786-792.
- [7] 刘春丽, 陈爽. 科学文献中的知识实体抽取与评价研究综述[J]. 现代情报, 2023, 43(12): 143-163.
- [8] 杨乐乐, 龙海, 姚克宇, 等. 基于本体推理和语义网检索的中西药相互作用知识发现研究[J]. 中华中医药学刊, 2024, 42(9): 13-17.
- [9] 郭利敏, 付雅明. 以大语言模型构建智慧图书馆: 框架和未来[J]. 图书馆杂志, 2023, 42(11): 22-30.
- [10] Bill Gates. AI is about to completely change how you use computers[EB/OL]. (2023-12-31) [2024-10-20]. <https://www.gatesnotes.com/AI-agents>.
- [11] 史永恒. 基于开源大模型的短视频与直播 AI 实训系统的设计与实现[J]. 信息技术与信息化, 2024(8): 195-198.
- [12] 赵一凡, 卞良, 丛昕. 数据清洗方法研究综述[J]. 软件导刊, 2017, 16(12): 222-224.
- [13] 易寒冰, 刘倩. 一种基于自动标注语料的热点事件情感分析方法及应用[J]. 数据与计算发展前沿, 2022, 4(5): 129-137.
- [14] 程顺航, 李志华, 魏涛. 融合自举与语义角色标注的威胁情报实体关系抽取方法[J]. 计算机应用, 2023, 43(5): 1445-1453.
- [15] 宋纯花. 基于人工智能的神经网络优化英语机器翻译[J]. 现代电子技术, 2024, 47(3): 80-84.
- [16] 丽平, 王恒山. 枣于多 Agent 的中文专利本体自动化构建方法[J]. 情报理论与实践, 2012, 35(3): 109-114.
- [17] 王笑尘, 张坤, 张鹏. 多视角看大模型安全及实践[J]. 计算机研究与发展, 2024, 61(5): 1104-1112.
- [18] 孙雨生, 曾俊皓. 向量数据库及其应用研究[J]. 科技情报研究, 2024, 6(4): 11-24.
- [19] 唐昆, 李白杨, 张心源. 基于主客观融合的人工智能跨模态生成内容质量及效能测度研究[J]. 情报理论与实践, 2024, 47(11): 150-161.

Development and Application of an AIGC-based Intelligent Management Assistant for Libraries: Taking the Application of AIGC in Library Knowledge Services as an Example

ZHANG Bing, HU Tingting, YU Quan

(Guangxi Zhuang Autonomous Region Science and Technology Information Research Institute, Nanning 530022, China)

Abstract: The methods, tools, technical frameworks, and application practices for knowledge mining were explored using AIGC (artificial intelligence generated content) in combination with the RAG (search enhanced generation) and stable diffusion models, as well as AI Agents, in order to development and application an AIGC-based intelligent management assistant for libraries, contributing to industrial development and knowledge dissemination. The research focuses on the development process, model principles, and utilization of AIGC technology to expand library knowledge services of the AIGC-based intelligent management assistant for libraries. Furthermore, the demands of industrial development and envisions future service expansion capabilities were integrated. Through the development and practice of this project, an efficient, intelligent, and personalized intelligent management assistant for libraries has been constructed. The innovative application of AIGC technology in library knowledge services was explored, aiming to provide some references for the specialized and scenario-based application of large models in the library and information science field.

Keywords: AIGC(artificial intelligence generated content); library; intelligent management assistant; knowledge service