



图书情报工作
Library and Information Service
ISSN 0252-3116, CN 11-1541/G2

《图书情报工作》网络首发论文

题目: 检索增强思维在图书馆知识挖掘服务中的应用研究
作者: 郭利敏, 付雅明, 刘悦如
收稿日期: 2024-10-15
网络首发日期: 2025-04-01
引用格式: 郭利敏, 付雅明, 刘悦如. 检索增强思维在图书馆知识挖掘服务中的应用研究[J/OL]. 图书情报工作.
<https://link.cnki.net/urlid/11.1541.G2.20250401.1354.002>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

检索增强思维在图书馆知识挖掘服务中的应用研究

■ 郭利敏¹ 付雅明² 刘悦如³

¹ 上海图书馆 / 上海科学技术情报研究所 上海 200031

² 上海大学文化遗产与信息管理学院 上海 200444

³ 同济大学图书馆 上海 200092

摘要：【目的/意义】聚焦大语言模型时代图书馆知识挖掘服务的发展现状，探索结合传统检索方法和检索增强思维，为用户提供既深刻又逻辑严密的知识信息，旨在提升图书馆的知识服务水平。【方法/过程】采用实验研究方法，以上海图书馆人名规范库为数据源，重点挖掘与宋代茶文化相关的信息，目标是从馆藏数据中揭示更多关于这一时期茶文化的细节，为用户提供更加全面和深入的视角。【结果/结论】实验结果表明，检索增强思维方法在图书馆知识挖掘服务中具有一定优势。该方法不仅有助于提高数据挖掘的效率和准确性，还能为图书馆知识挖掘服务提供新的技术解决方案，对智慧图书馆的建设具有参考价值。

关键词：知识挖掘 检索增强思维 智慧图书馆服务 大语言模型

分类号：G250

引用本文：郭利敏，付雅明，刘悦如. 检索增强思维在图书馆知识挖掘服务中的应用研究[J]. 图书情报工作, 2025, 69(X): 00-00. (Citation: Guo Limin, Fu Yaming, Liu Yueru. Research on the Application of Retrieval Augmented Thoughts in Library Knowledge Mining Services[J]. Library and Information Service, 2025, 69(X): 00-00.)

1 引言 / Introduction

随着人工智能技术的迅猛发展，人类数字化社区规模逐渐扩大，数据和信息在人类社会、物理空间和网络空间往更加深入的方向融合，人工智能 2.0 时代已然到来^[1]。人工智能技术已经不仅仅是简单的自动化工具，而是成为认知计算的核心，能够理解、学习、推理并与人类共同协作。人工智能技术的迅猛发展在加速信息产生速度与规模的同时，也改变了知识的获取、处理与分享方式^[4]。有学者表示即将到来的人工智能生成内容 (AI-generated content, AIGC) 时代将根本性地变革计算模式，把当前以数据驱动的模式转换为结合数据驱动与知识引导的模式，从而支持多样化的人类决策^[5]。

AI2.0 时代，图书馆作为知识发现、组织、传播和保存的专业机构，身处从数据时代到计算时代

所带来的知识内容、学术交流、知识资源、知识环境的“新常态”下，其自身与所提供的服务都需要“重新定位”，从静态独立转为一种“基于知识内容与知识分析计算的创新孵化器”^[2]。而在“新常态”下，用户及其所身处的阅读、学习、研究等场景也发生了变化，互动式研究性学习、开放科学研究等多种新的模式与相关活动对图书馆的信息服务提出了新的要求，构建并提供可计算的知识资源、工具与知识挖掘路径成为可行之策^[2]。其中知识挖掘作为知识发现与服务中的核心环节与方法，可以借助人工智能技术得到效能提升，以满足“新常态”下对于知识挖掘的新需求。知识挖掘是从海量数据中提取有价值的知识的过程，是数据挖掘的衍生概念。与传统的数据挖掘不同，知识挖掘不仅限于发现隐藏在数据中的模式和信息，还包括对这些信息的再加工、提炼和应用，从而生成具有实际意义和决策

* 本文系上海市白玉兰人才计划浦江项目“基于生成式人工智能的文化遗产知识服务研究”（项目编号：24PJC036）和上海市哲学社会科学项目“智慧数据驱动的大都市图书馆科研创新支撑服务”（项目编号：2020BTQ005）研究成果之一

作者简介：郭利敏，副研究馆员，硕士；付雅明，助理研究员，博士，通信作者，E-mail: ymingfu@126.com；刘悦如，副研究馆员，硕士。

收稿日期：2024-10-15 修回日期：2024-12-17

版权所有 ©《图书情报工作》杂志社有限公司，未经许可不得转载 (Copyrights © LIS Press Co., Ltd. Reproduction is prohibited without permission)

价值的知识。其概念包含了数据发现、知识发现、信息发现、数据开采和智能数据分析等多种环节与形式^[3]。

AI2.0 时代,图书馆的知识挖掘能力与服务面临着前技术赋能的机遇和挖掘深度的挑战,一方面,AIGC 技术的快速发展为图书馆提供了更强大的数据分析和知识发现工具,在信息资源的组织与检索方面变得更加高效精确;另一方面,图书馆的知识挖掘也面临知识内容版权保护、数据准确性、技术模型适配度等问题^[7]。而图书馆的知识挖掘也与图书馆文献分析、知识服务、知识发现等领域息息相关,可以说探索研究 AI2.0 时代的图书馆知识挖掘服务是在新的信息与技术环境下的重要课题,也是创新图书馆知识服务的关键要点。

检索增强思维 (retrieval augmented thoughts, RAT) 是一种通过信息检索迭代修订思维链,提升大语言模型在复杂任务中推理和生成能力的系统模式^[8]。在知识挖掘领域,这种模式不仅能够理解和处理输入的信息,还能够围绕输入的信息进行主动思考,结合相应的数据源,进一步迭代修订生成的内容,为用户挖掘更为准确和隐藏的深层知识。

本文通过实验研究,探讨检索增强思维方法在图书馆知识挖掘服务中的应用潜力。实验以上海图书馆

开放人名规范库为数据对象,以“宋代茶文化的‘龙凤盛世’”为主题,旨在通过检索增强思维技术,深入挖掘与宋代茶文化相关的历史人物、事件和文化现象,探索 RAT 在图书馆知识挖掘服务中的应用实现路径。

2 大语言模型时代的图书馆知识挖掘服务 / Library Knowledge Mining Services in the LLM Age

在 AIGC 时代,图书馆知识挖掘服务往往被看作为智慧图书馆服务的重要组成部分^[9],知识服务通过有机组织数字资源和深度理解知识结构,利用专业工具协助用户进行知识挖掘、计算、试验和评价^[10]。而知识挖掘则是利用数据挖掘、自然语言处理、知识图谱构建等技术,从海量数据中自动提取、分析与揭示隐藏的知识、新模式和关联关系^[11]。数据挖掘是知识挖掘的基础,通过统计分析、机器学习和模式识别等手段从大量数据中获取可用信息^[12]。知识挖掘则在数据挖掘结果的基础上进行深层次分析,旨在揭示数据背后的隐性知识与价值。

总的来说,知识服务、数据挖掘和知识挖掘 3 者之间的关系,可以理解为一个从基础到高阶的过程,如表 1 所示。

表 1 数据挖掘、知识挖掘与知识服务的定义、特点与技术工具对比表

Table 1 Definitions, characteristics, and comparison of technical tools for data mining, knowledge mining, and knowledge services

概念	数据挖掘	知识挖掘	知识服务
定义	从大量数据中提取有用信息和模式	从数据和信息中提取有价值的知识	提供基于知识的解决方案和服务
研究对象	结构化和非结构化数据	结构化数据、非结构化数据、背景信息等	用户需求、知识库、信息资源等
目标	发现数据中的隐藏模式和关联	从杂乱的数据中挖掘知识内容,提供深层次的知识理解	提供符合用户需求的知识服务
一般使用方法	统计分析、机器学习、分类、聚类	自然语言处理、知识图谱、语义分析	信息检索、智能推荐
技术工具	数据库、数据仓库、统计软件	自然语言处理工具、知识图谱工具、人工智能工具	知识管理系统、推荐系统

知识挖掘在图书馆中的应用可以极大地提升图书馆的服务能力和面向用户的精准服务水平^[13]。一方面,对馆藏资源进行深度分析,促进知识重组与价值凸显;另一方面,通过用户行为数据分析,更精准地理解用户偏好与需求,提供个性化知识推荐,优化馆藏结构和资源利用。

知识挖掘和大语言模型之间存在密切的关系,尤其在数据处理、知识发现和知识库更新等方面相辅相成。目前在知识挖掘领域,学者们^[14]设计了生成式情报学术语抽取方法,并进而实现了基于术语发现于多维度知识挖掘;探索了 GPT-4 及 Google Bard 模型对提取于挖掘文本内容、挖掘抽象概念等方面的

能力^[15];也有运用检索增强生成 (retrieval-augmented generation, RAG) 和大语言模型微调在特定知识领域进行针对检索任务的内容挖掘与优化^[16]。大语言模型具备强大的自然语言处理、生成与推理能力,可有效处理非结构化文本数据,促进隐藏知识模式的发现与更新。在数据处理和分析方面,大语言模型能够有效处理非结构化文本数据,提升知识挖掘的准确性和效率^[17];在知识发现和生成方面,大语言模型可以通过对大量数据的学习,发现隐含的知识模式并生成新的知识^[18];在知识库更新和维护方面,大语言模型通过不断学习最新数据,更新和维护知识库的内容,保持其时效性和准确性^[19]。

AIGC 时代的快速发展,促使图书馆知识服务走向智能化、语义化与细粒度化。借助数字知识表示、语义关联、深度学习与认知计算等新一代技术,图书馆正从传统的资源提供者向知识挖掘与发现的创新主体转型,构建以知识计算和知识发现为核心的新一代智慧知识服务体系^[20-21]。

3 基于检索增强思维的图书馆知识挖掘服务 /RAT-based library knowledge mining services

3.1 基于检索增强思维的知识挖掘

大语言模型在处理多步推理任务时,如代码生成、数学推理等,需要依据上下文信息进行准确推理。然而,由于缺乏对真实世界知识的直接访问,模型可能会生成与事实不符的中间推理步骤,导致最终结果的不准确,即所谓的幻觉问题^[21]。为了解决这一问题,研究者们希望结合信息检索技术,通过检索与任务相关的信息来辅助模型进行更准确的推理。在检索和推理方面分别诞生了 RAG 和思维链(chain of thoughts, CoT)两种关键技术。

RAG^[22]是一种结合了检索和生成的模型架构,旨在提高生成式模型在问答和对话任务中的性能。RAG 模型结合了检索式问答和生成式问答的优点,通过在生成式模型中集成检索式模型,使其能够在生成答案时引入外部知识和信息。这种结合可以帮助生成式模型更好地理解问题并生成更准确和丰富的答案。RAG 模型已被证明在多项自然语言处理任务中取得了显著的性能提升。但是对于需要多步骤推理的复杂任务而言,再将任务转化成有效的搜索查询上存在一定的挑战,随着步骤的增加,上下文内容的增长,导致从内容中提取检索要素的准确性会逐步降低,从而限制了 RAG 的适用性^[23]。

CoT 技术是一种旨在增强语言模型在需要复杂推理的任务上表现的方法^[24]。它利用大语言模型的逻辑推理能力,根据输入的文本,生成连贯的、具有逻辑性的输出文本。可以用于各种自然语言处理任务,如文本生成、对话系统等。通过学习大量的文本数据,其能够理解语言的语法结构和语义含义,从而生成具有连贯性和合理性的文本内容。通过鼓励语言模型首先输出中间推理步骤,然后再总结为最终答案,来模拟人类的分步推理过程。这种方法已经在一些数学推理、代码生成等任务中取得了显著的性能提升。由于没有对中间推理步骤进行直接监督,语言模型可

能会由于缺乏相关领域知识或受到幻觉的影响而产生错误^[25]。

RAT 则是融合了 RAG 与 CoT 两者的优点,利用 RAG 从可信数据源提取相关文本来辅助模型生成,CoT 提示则模仿人类解决问题的思维过程,鼓励模型分步骤地思考和推理,使得模型生成的回答能够更好地进行上下文感知,减少幻觉,提高事实准确性。作为一种协同了 CoT 和 RAG 的 AI 提示策略,检索增强思维有助于解决长任务推理和长文本生成任务,对于代码生成任务(HumanEval 和 HumanEval+ 的基准测试)、数学推理任务(GSM8K 和 GSMHard 数据集)准确率有显著提升,在 Minecraft 任务规划中生成计划的可执行性和可行度方面均优于基准线方法,在创造性写作任务中,检索增强思维的文本质量、准确性和创造性方面得到了人类评估者的高度评价^[8]。

因此,基于 RAT 的知识挖掘服务能够让大语言模型自主地从多个信息源中提取数据,提供全面多样化的信息;能够更好地理解文本语义,推测用户的意图以及对检索结果的内容整合,能够极大程度提升知识挖掘的有效性和效率。相较于 RAG 通过信息检索来提高生成模型的回答质量,RAT 侧重于通过逐步推理和信息修正来提高长任务的推理能力。

3.2 基于检索增强思维的图书馆知识挖掘服务流程框架设计

基于检索增强思维的图书馆知识挖掘服务流程是一种创新的信息服务模式,该流程主要分为三个核心步骤:思考、检索和汇总,如图 1 所示。思考步骤是整个知识挖掘服务流程的起点,用户需要对自己的研究主题、兴趣点或问题进行深入思考,以确定需要检索的具体信息类型。例如,用户可能需要寻找关于某个历史事件的详细资料,或者需要了解某个科学领域的最新研究成果。在这一阶段,用户可以通过与图书馆工作人员的交流,或者利用图书馆提供的在线咨询工具,来帮助自己更准确地界定信息需求。

思考步骤需要利用到大语言模型的自然语言理解能力,根据用户的输入分析出用户的需求(包括潜在的需求)和意图,并根据意图模拟人的思考。把输入 I 拆分成若干个相关的思考步骤,其中 n 的多少由大语言模型根据输入自行决定。

检索步骤是整个服务流程的核心。在这一步骤中,大模型结合用户的需求,从数据库和其他数据源中检索相关信息。在思考步骤的基础上,围绕每一个思考步骤的主题,在内容集合 C 中检索与主题最相

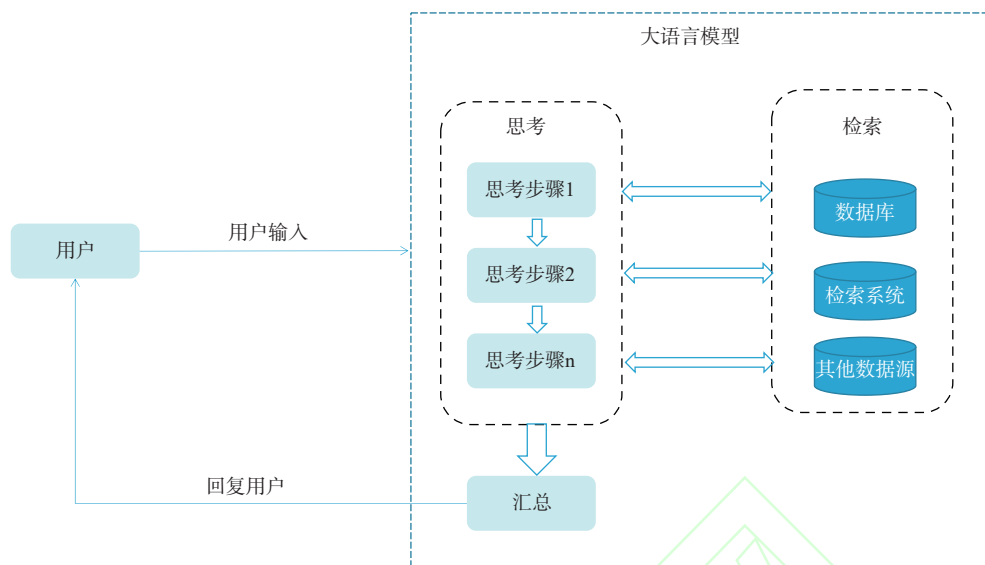


图 1 基于检索增强思维的图书馆知识挖掘服务框架
Figure 1 Library knowledge mining service framework based on RAT

关的条内容，内容集合 C 可以是数据库，也可以是第三方 API，包含结构和非结构化数据。这一步骤的关键在于实现信息的快速定位和精确匹配，以满足用户的个性化需求。

汇总则是将所有检索结果与各自的思考步骤相结合，将检索到的信息进行整合和优化，生成相应的段落内容，最后将所有的段落内容与用户的输入 I 整合起来生成最终的挖掘内容返回给用户。在这一阶段，知识挖掘系统会对检索结果进行筛选和排序，去除重复和不相关的信息，保留最有价值的内容。此外，汇总步骤还可能包括对信息的进一步加工，如摘要编写、关键点提炼等，以帮助用户快速把握信息的核心要点。

基于检索增强思维的图书馆知识挖掘服务流程是一种以用户为中心，以技术为支撑，以知识创新为目标的信息服务模式。在整个服务流程中，检索增强思维的概念贯穿始终，检索增强思维不仅关注信息的检索效率，更注重提升用户的信息处理能力。它通过思考、检索和汇总 3 个步骤，实现了信息的快速定位、精确匹配和有效整合，为用户提供了一种知识来源广泛同时高度提炼的知识获取能力和知识“深潜”体验。

4 实验与效果分析 /Experiment and effect analysis

本研究的实验内容以“宋代茶文化的‘龙凤盛世’”为主题，以知识挖掘为目的探索宋代茶文化的繁荣和发展历程。实验运用大语言模型的分析能力，结合上海图书馆开放人名规范库中的规范化知识内

容，挖掘与宋代茶文化相关的历史人物、事件和文化现象，旨在从用户视角探索增强检索技术对知识挖掘的价值，验证 RAT 应用在知识挖掘服务中的可行性。具体实验流程如图 2 所示，其过程可类比于学术文章的撰写，即首先围绕主题构建文章框架，然后逐一撰写各段落。

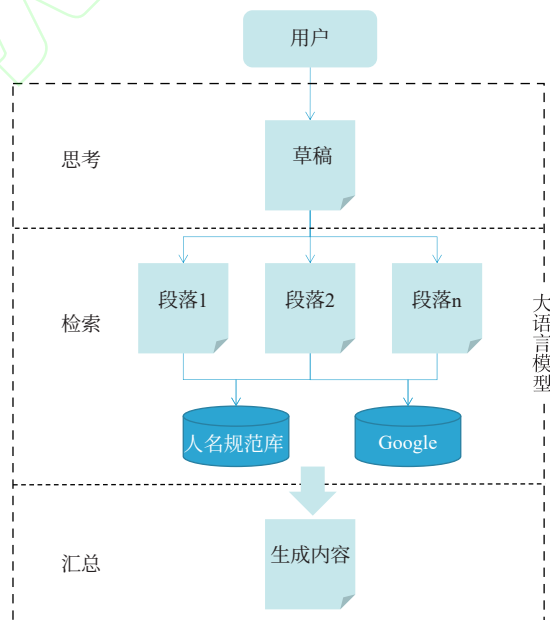


图 2 基于检索增强思维的图书馆知识挖掘实验流程
Figure 2 Experimental process of library knowledge mining based on RAT

在思考阶段采用 CoT 方法，通过提示词引导模型逐步展开思考，以系统性分析主题。分析结果按照特殊分隔符“\n\n”进行分段，形成初步草稿。在此阶段，可以指定模型从特定角度进行思考，但在本实验中未作特别指定，允许模型自主选择其认为合适的

角度进行讨论，以期望模型能够挖掘出不一样的知识。每个角度均独立成段，且每段内容均明确了其核心论断（即段落的中心思想），具体草稿样例如图 3 所示。

宋代是中国茶文化发展的黄金时期，具体体现在以下几个方面：
第一，宋朝的茶文化与众多历史人物密不可分，根据人名库里的数据，与茶文化相关的人物在宋朝尤其众多。[...]（具体说明）
第二，宋朝的各个阶层均有饮茶的习惯。这不仅仅是上层贵族的雅好，茶文化已经渗透到普通百姓的日常生活。[...]（具体说明）
第三，宋代与茶有关的作品众多。[...]（具体说明）
综上所述，宋代的茶文化不仅在社会的各个层面得到了广泛的普及，并且在文学和艺术作品中留下了浓墨重彩的一笔，成为中国历史上茶文化的重要一章。

图 3 检索增强思维草稿部分示例
Figure 3 Example of draft section for RAT

进入检索阶段，实验采用 RAG 策略，针对每个段落进行详细撰写。首先，通过提示词提炼段落中心思想的检索关键词，并使用 Google 检索 API 获取排序前 3 的相关文献（包括 URL、简介和标题）。随后，以后通过提示词的形式让模型对文献资源进行命名实体识别，提取段落中出现的人名，并通过接口在人名规范库中检索相关详细信息（如人物详情页），作为本段落的补充文献，具体检索示例如表 2 所示。最终，结合检索结果与段落中心思想，对各段落进行深入撰写。

在汇总阶段，本研究将各个独立段落的内容进行整合，形成连贯的知识总结，如图 4 所示。最后，对生成的内容进行校对，确保其内容与主题的相关性。

宋代是中国茶文化发展的黄金时期，具体体现在以下几个方面：
第一，宋朝的茶文化与众多历史人物密不可分，根据人名库里的数据，与茶文化相关的人物在宋朝尤其众多。例如，宋徽宗赵佶不仅是一位帝王，更是茶文化的热衷推崇者，他在《大观茶论》中详细论述了茶叶的制作和饮用方法。此外，还有朱熹、苏轼等文化名人，他们不仅是儒家学者和文人，也是茶文化的积极参与者和推广者。
第二，宋朝的各个阶层均有饮茶的习惯。这不仅仅是上层贵族的雅好，茶文化已经渗透到普通百姓的日常生活。上层社会有精细的茶宴和品茶活动，文人雅士则常在诗中咏茶、品茶，而百姓们则在日常生活中饮用茶水以消暑解渴。
第三，宋代与茶有关的作品众多。宋徽宗的《大观茶论》、蔡襄的《茶录》、欧阳修的《归田录》都详尽描述了茶的制作、品鉴和社会功能。此外，苏轼的诗词中也有大量关于茶的内容，例如“从来佳茗似佳人”等作品，体现了他对茶的热爱。
综上所述，宋代的茶文化不仅在社会的各个层面得到了广泛的普及，并且在文学和艺术作品中留下了浓墨重彩的一笔，成为中国历史上茶文化的重要一章。

图 4 总结阶段的临时内容
Figure 4 Temporary content during the summarization phase

4.1 实验模型选型和数据选择

由于实验环境中 GPU 计算资源有限，本文以 ChatGPT3.5-Turbo-16k、ChatGPT4-32k、ChatGPT-4o 模型作为实验模型，通过接口可以切换成其他提供服务的模型。实验数据上，本文选择上海图书馆人名规范库（<https://names.library.sh.cn/mrgf/home/index>）作为内部数据源。由于系统输入仅为一句话，且并无其他资源系统为实验提供相关的数据资源，因此选择 Google 检索系统作为实验的外部数据源。

4.2 试验方法和评估指标

本实验选用的模型以直接回答和运用 RAT 技术两种方法，对“宋代茶文化的‘龙凤盛世’”这句输入进行内容生成，并对生成的内容进行评估，具体两种方法的特点描述如表 3 所示。

由于本实验的主题为开放性问题，并没有一个标准的答案。因此，实验选择以人工评价的方式从历史人物和相关作品、逻辑性与系统化组织能力、文化现

象的解读、写作和表达与资料来源和引文 5 个方面对实验结果进行综合评估，并用不同颜色在实验结果中对不同指标的反映情况进行标注，具体的评估指标与对应评估目的如表 4 所示。

表 2 Google 和人名规范档检索结果示例
Table 2 Example of Google search and names of Shanghai library search result

查询源	查询关键词	查询结果
Google	宋代茶文化 赵佶 苏轼 朱熹 茶作 品	{ 'title': '《武夷山茶文化》之四《茶风古韵》解说词 - 武夷山奇苑茶业有限公司', 'link': 'https://www.wuyishantea.com/detail.asp?id=1363', 'snippet': 'Jul 8, 2009 ... 苏轼在《咏茶》诗中的描写：“君不见武夷溪边粟粒 ... 茶文化的重要文献，对研究宋代茶事有着重要的价值。...” 朱熹以茶喻学，认为学问要像茶一样，贵在真味 \xa0...'} }
人名规范库	苏轼	http://data.library.sh.cn/entity/person/fia6wo2nwkdcrl

表 3 RAT 直接回答两种方式的比较
Table 3 Comparison of two methods for direct answering in RAT

方法	描述
直接回答	直接使用问题本身作为检索查询，完全依赖模型的训练数据的内容
RAT	结合 RAG 和 CoT 的优势，通过动态调整检索查询来增强推理能力； 优点是可以充分利用模型的推理能力和外部知识，缺点是实现相对复杂

依据上述的试验方法和评估指标，我们对模型进行了两个对照实验。实验 1 用以比较不同模型使用 RAT 方法的效果，确定不同模型的效果对比。实验 2 则是比较同一模型在直接回答和使用 RAT 方法后回答的效果，以确定 RAT 方法的有效性。

表 4 试验评估指标及目的

Table 4 Experimental evaluation metrics and objectives

评估指标	评估目的	标注颜色
历史人物和相关作品	考察研究内容是否符合主题,且是否能够结合本地知识库	红色
逻辑性与系统化组织能力	研究和思考问题的逻辑性和系统性,是否能形成清晰连贯的论述和观点	绿色
文化现象的解读	是否深入理解文化的多样性和复杂性,以及文化如何影响个体和社会	蓝色
写作和表达	是否包括清晰地传达观点、构建合理的论证结构、使用准确的语言和语法	紫色
资料来源和引文	是否正确引用本地知识库,并进行规范著录	褐色

4.3 实验结果分析

实验中关于引文部分 [citation:x] 来表示引文位置,其中 x 表示第 i 个引文。另外对于人名规范档的引用,实验以展示文档的对象表示符 OID 为方式进行引用,如“陆游”的 OID 为 <http://data.library.sh.cn/entity/person/hz8kvfmism0v0wz8>。

实验 1 设计使用 RAT (主题挖掘工具) 对

ChatGPT-3.5、ChatGPT-4 和 ChatGPT-4o 进行分析,以探讨它们对输入主题“宋代茶文化的‘龙凤盛世’”的挖掘能力。通过这种分析,希望能够评估不同模型在表 3 中所列评估框架下的表现和效率。具体而言,该实验旨在检测 3 种不同版本的 ChatGPT 模型在处理特定文化主题挖掘时的各自的优劣。实验 2 则设计用于评估同一模型在使用和不使用 RAT 情况下的表现差异。通过观察和比较模型对于同一主题挖掘的输出结果,来评判 RAT 对模型表现的影响。

通过上述两个实验,不仅能深入了解不同版本的 ChatGPT 在处理知识主题挖掘时的表现差异,还能评估 RAT 在知识挖掘任务中的实用价值。这些分析和比较为进一步优化模型提供了重要的参考依据,并为相关研究领域的实践提供了翔实的数据支持。

从实验 1 的结果而言,不同模型使用 RAT 进行知识挖掘的结果分如下:

茶文化在宋代得到了广泛的传承与发展,因此宋朝被称为“茶盛时代”[citation:4]。从文献中可以得知,茶在宋代的社会生活中占据重要地位,不论是在皇宫还是普通人家,人们都有饮茶的习惯。茶文化在宋代涵盖了各个阶层的人群。

首先,在皇宫中,茶作为一种奢侈品,成为宫廷贵族喜爱的饮品。宋徽宗在位期间,御茶园开始尝试创新,除了以往的传统贡品,还精制的许多新品种的贡茶,供皇帝品尝。这些贡茶包括白茶、龙园胜雪、御苑玉芽、万寿龙芽等[citation:5]。

其次,在文人雅士中,茶也有浓厚的兴趣,并将其作为创作灵感的源泉。以诗人陆游为例,他是南宋时期的一位爱国诗人,其诗作中常有对茶的描写[citation:1]。他号称“放翁”,将茶喻为自己的伴侣,表达了对茶文化的热爱。

此外,在宋代的社交生活中,茶馆成为了重要的场所。人们在茶馆中品茶、交流和欣赏艺术表演。茶馆老板娘赵盼儿在《梦华录》中展示了点茶高手技艺,以及斗茶和茶百戏等茶艺表演,使人们更加了解了宋代茶文化的繁荣和茶艺的精湛。

茶文化还在宋代涌现了许多与茶相关的作品。茶经、茶谱等专门介绍茶的著作在宋代广泛流传。例如,王安石的《议茶法》中阐述了茶的重要性,对当时的茶文化有着重要的影响[citation:3]。词人李清照在《临安春雨初霁》和《满庭芳》等作品中也有对茶的描写[citation:2]。

综上所述,茶文化在宋代的“龙凤盛世”中得到了广泛传承与发展。茶在不同阶层的人中间都有一定的地位,从皇宫的贵族到文人雅士,茶都是重要的饮品。茶馆成为了社交活动的重要场所,人们在茶馆中品茶、交流和欣赏表演。许多作品也与茶文化密切相关,在文学、艺术和社会生活中发挥了重要的作用。

[citation:1]] 陆游 (<http://data.library.sh.cn/entity/person/hz8kvfmism0v0wz8>)
[citation:2]] 李清照 (<http://data.library.sh.cn/entity/person/e5mfjnv4257kg62z>)
[citation:3]] 王安石 (<http://data.library.sh.cn/entity/person/4348dwg8wnm0hail>)
[citation:4]] 宋朝 (<http://data.library.sh.cn/entity/person/rsxxjgz7cs1mgkt>)
[citation:5]] 白茶 (<http://data.library.sh.cn/entity/person/rzpenxmzyv5sap7d>)

图 5 ChatGPT3.5 使用 RAT 的实验结果

Figure 5 Result of RAT using ChatGPT3.5

如图 5 所示,ChatGPT3.5 在历史人物和相关作品方面,结合人名规范库挖掘出白茶、宋朝、李清照、陆游和王安石这 5 个历史人物,其中白茶和宋朝这两个人名存在于规范库中但与宋朝和茶文化无关,属于错误的挖掘;对于作品《临安春雨初霁》和《满庭芳》也属于错误挖掘,前者作者为陆游,后者作为词仅列出了词牌名缺少对应的标题,无法判断作者。逻辑性与系统化组织能力方面,笼统地介绍了茶文化在不同社会阶层中的地位和影响。文化现象方面,指出文人雅士在日常生活中的对茶的兴趣,结合《梦华录》(当代关于北宋中叶的电视剧)阐述了茶在百姓中的地位,这个挖掘需要根据实际情况具体分析其价值性。写作和表达方面,写作清晰、句子流畅、表达准确,采用“首先、其次”的分段

结构化表达。引文方面,能够在具体位置插入相关的引文,但并没有对引文与主题的相关性进行筛选,存在一定的错误。

如图 6 所示 ChatGPT4 在历史人物和相关作品方面,结合人名规范库挖掘出蔡襄、文天祥、陆羽、朱熹、范宽、郭熙和陆游这 7 个历史人物,遗漏了赵恒;对于作品《卜算子·咏梅》为陆游一首咏梅词,“功盖三分国,名成八阵图……”为杜甫的《八阵图》,与茶无关,《朱子家礼》是朱熹的礼学著作,不知与茶有什么关系,均属于模型幻觉。逻辑性与系统化组织能力方面,从皇家、文人墨客和百姓 3 方面介绍了社会各阶层的饮茶习俗;从诗词、书籍和绘画 3 方面介绍了与茶文化有关的历史人物及其作品。文化现象方面,指出皇家是茶文化的重要推动者,讲究茶

艺的精致与雅致；文人雅士喜欢边品茶边吟诗作画，将茶与文化艺术融为一体；茶馆是百姓社交的重要场所。写作和表达方面，文章写作清晰流畅，采用段落

符号的形式分类描述，便于阅读。引文方面，能够在具体位置插入相关的引文，但并没有对引文与主题的相关性进行甄别，存在一定的错误。

宋朝在中国历史上被誉为“龙凤盛世”，在这个时期，茶文化发展到了极致。早在唐宋之交，茶已成为日常生活中不可分割的部分，到了宋代，整个社会的文化氛围推动茶的消费和发展，形成了“尚茶”的社会风尚。尤其以北宋为代表，茶文化达到了崭新的高度。以下将就宋代不同阶层的饮茶习俗，以及茶文化相关的作品进行介绍。

1. 社会各阶层的饮茶习俗：

(1) 皇家：皇家是茶文化中的重要推动者。皇帝为了彰显其威严与品位，讲究茶艺的精致与雅致，尤其偏好贡茶。北宋皇帝赵恒就是一个深爱茶文化的君主，他好读书喜爱茶，不仅自己饮茶，还经常赠与臣子。

(2) 文人墨客：宋朝讲究文人雅致，因此，文人墨客对茶艺有着独特的热爱。他们喜欢边品茶边吟诗作画，将茶与文化艺术融为一体。如经学家朱熹，他在《朱子家礼》中记载了茶的采摘、炒制、研磨等详细过程，展现了他对茶艺的深入理解和独特见解[[citation:4]]。

(3) 百姓：普通百姓也很喜欢饮茶，茶已经成为他们日常生活的一部分。茶馆是百姓社交的重要场所，他们在这里饮茶、聊天、娱乐，展现了宋朝茶文化的普及性。

2. 茶文化相关的作品：

(1) 书籍：北宋官员蔡襄的《茶录》、陆羽的《茶经》以及朱熹的《朱子家礼》等，这些书籍展示了宋代茶文化的卓越与深远影响[[citation:1]][[citation:3]][[citation:4]]。

(2) 诗词：宋代有许多诗人在诗词之中表现了对茶的热爱，如陆游的《卜算子·咏梅》[[citation:7]]，“功盖三分国，名成八阵图。江流石不转，遗恨失吞吴。茶余饭饱诗书足，乘月裴回顾挟飞上层。”文天祥等人的诗词中也常常融入对茶的描绘[[citation:2]]。

(3) 绘画：宋代的绘画大师如郭熙、范宽等人的画作中，也可以看到茶文化的影子[[citation:5]][[citation:6]]。他们的画作中经常会出现品茶的场景，画面雅致温馨，生动展示了宋代的茶饮生活。

总之，宋代是中国茶文化的鼎盛时期，茶叶已经融入了社会各阶层生活之中，形成了深远而广泛的影响。无论是皇家、文人还是普通百姓，他们都用各自独特的方式表现了对茶的喜爱。书籍、诗词和绘画等艺术形式，也都反映了当时深厚的茶文化底蕴。

[[citation:1]] 蔡襄 (http://data.library.sh.cn/entity/person/b2vflq1z4e3qzm3)
[[citation:2]] 文天祥 (http://data.library.sh.cn/entity/person/7wg0wfrmf63nk2h8)
[[citation:3]] 陆羽 (http://data.library.sh.cn/entity/person/gwof2rfu6nklz4oc)
[[citation:4]] 朱熹 (http://data.library.sh.cn/entity/person/r5vmxq29ki9gdjnj)
[[citation:5]] 范宽 (http://data.library.sh.cn/entity/person/z2ryulo8gin8rxlo)
[[citation:6]] 郭熙 (http://data.library.sh.cn/entity/person/72bfxoyn1pslfzlj)
[[citation:7]] 陆游 (http://data.library.sh.cn/entity/person/hz8kvfmism0v0wz8)

图 6 ChatGPT4 使用 RAT 的实验结果
Figure 6 Result of RAT using ChatGPT4

```
## 宋代茶文化的“龙凤盛世”
宋代是中国茶文化发展的黄金时期，具体体现在以下几个方面：
### 历史人物与茶文化
宋朝的茶文化与众多历史人物密不可分，根据人名库里的数据，与茶文化相关的人物在宋朝尤其众多。例如，宋徽宗赵佶不仅是一位帝王，更是茶文化的热衷推崇者，他在《大观茶论》中详细论述了茶叶的制作和饮用方法。
当时的新制贡茶被进献到宫廷，徽宗皇帝与群臣品饮，并在画作《文会图》中描绘了当时的盛景。此外，著名的宋代茶文化推广者蔡襄，不仅在文学和书法上有着重要成就，更是在茶叶史上留下了自己的名字，他改进了龙团凤饼的外形和制作工艺，使得其品质极为珍贵[[citation:1]]。还有苏轼、欧阳修等文化名人，他们不仅是儒家学者和文人，也是茶文化的积极参与者和推广者[[citation:2]][[citation:3]]。
### 茶文化的广泛普及
宋朝的各个阶层均有饮茶的习惯。这不仅仅是上层贵族的雅好，茶文化已经渗透到普通百姓的日常生活。上层社会有精细的茶宴和品茶活动，文人雅士则常在诗中咏茶、品茶，而百姓们则在日常生活中饮用茶水以消暑解渴。蔡襄督造茶事时，还曾亲自到茶山观察茶叶生长，改进工艺，造出了极为精绝的小龙团，甚至连欧阳修都在诗中感叹其稀有珍贵[[citation:1]][[citation:2]]。
### 与茶有关的文学作品
宋代与茶有关的作品众多。宋徽宗的《大观茶论》、蔡襄的《茶录》被视为茶文化文献的顶峰之作，详尽描述了茶的制作、品鉴和社会功能[[citation:1]]。此外，苏轼的诗词中也有大量关于茶的内容，例如“从来佳茗似佳人”等作品，体现了他对茶的热爱[[citation:3]]。同时，欧阳修的《归田录》也提到了茶叶的稀有珍贵[[citation:2]]。
### 茶具文化
宋代还有丰富的茶具文化，尤其是建盏茶具。宋徽宗在点茶时常使用的兔毫盏，出自建窑，被称为建盏，这种茶具在日本和韩国被称为天目盏，至今仍被视为珍宝。著名的韩国陶瓷艺术家金俊因为对武夷山古窑址的喜爱，常年留在此地从事陶艺创作，展示了宋代制茶与茶具文化的深远影响。
综上所述，宋代的茶文化不仅在社会的各个层面得到了广泛的普及，并且在文学和艺术作品中留下了浓墨重彩的一笔，成为中国历史上茶文化的重要一章。
[[citation:1]] 蔡襄 (http://data.library.sh.cn/entity/person/b2vflq1z4e3qzm3)
[[citation:2]] 欧阳修 (http://data.library.sh.cn/entity/person/xrgok9phxhsun9z2)
[[citation:3]] 苏轼 (http://data.library.sh.cn/entity/person/fia6wo2nwkdcrrjl)
```

图 7 ChatGPT4o 使用 RAT 的实验结果
Figure 7 Result of RAT using ChatGPT4o

如图 7 所示 ChatGPT4o 在历史人物和相关作品方面，结合人名规范库挖掘出蔡襄、欧阳修和苏轼，遗漏了宋徽宗赵佶；相关作品在使用上均正确无误。逻辑性与系统化组织能力方面，分段逐步介绍了历史人物与茶文化的关系、茶文化在宋朝的普及程度、相关的文学作品和宋代的茶具，其中茶具是一个较为独特的挖掘点。文化现象方面，指出文人雅士常在诗中咏茶、品茶，而百姓们则在日常生活中饮用茶水以消暑解渴，此外还特别指出茶具建盏。写作和表达

方面，写作清晰流畅，采用 Markdown 格式进行表述，更适合于互联网阅读习惯。引文方面，能够在具体位置插入相关的引文。

综上所述，ChatGPT4 和 ChatGPT4o 系列模型在文本组织、分析和逻辑性上面优于 ChatGPT3.5（仅针对知识挖掘的内容进行文本阐述），并没有使用任何的带有符号形式的分类描述，相较于 GPT4 系列模型而言较为逊色。对于 GPT4 和 4o 而言，分别采用了结构化的表述形式（段落符号和 Markdown 格式），

表 5 各模型使用 RAT 的结果对比分析
Table 5 Comparative analysis of results from various models using RAT

评估指标	GPT3.5	GPT4	GPT4o
历史人物和相关作品(红色)	挖掘出人名陆游、李清照、王安石、宋朝、白茶,但宋朝和白茶与宋代茶文化无关,存在一定错误。作品方面存在一定错误	挖掘出蔡襄、文天祥、陆羽、朱熹、范宽、郭熙和陆游这七个历史人物,遗漏了赵恒。作品方面存在一定的错误	挖掘出蔡襄、金俊、欧阳修、苏轼符合主题,遗漏宋徽宗赵佶。作品方面无明显错误
逻辑性与系统化组织能力(绿色)	笼统地介绍了茶文化在不同社会阶层中的地位和影响	分别从皇家、文人墨客和百姓和诗词、书籍和绘画 6 方面介绍了与茶文化有关的历史人物及其作品	分析了相关历史人物及其作品,此外还挖掘出建盏茶具
文化现象的解读(蓝色)	仅指出文人雅士在日常生活中的对茶的兴趣	指出皇帝讲究茶艺的精致与雅致,尤其偏好贡茶;文人雅士喜欢边品茶边吟诗作画,将茶与文化艺术融为一体;茶馆是百姓社交的重要场所	指出文人雅士常在诗中咏茶、品茶,而百姓们则在日常生活中饮用茶水以消暑解渴,此外还特别指出茶具建盏
写作和表达(紫色)	具有可读性和自然性	提供明确的分级表达	写作清晰流畅,采用 Markdown 格式简洁易读,更适合于互联网阅读习惯
资料来源和引文(褐色)	存在错误引用	存在遗漏,相关作品存在错误	存在遗漏

宋代是中国茶文化发展的重要时期,众多与茶文化相关的人物在这个时代涌现。以下是一些与宋代茶文化相关的人物和作品:

1. 陆羽: 陆羽是茶叶栽培和制作技术的奠基人。他创作了《茶经》,成为了中国茶文化的经典著作。
2. 苏东坡: 苏东坡是宋代文学家、茶文化爱好者。他曾写下了许多有关茶的诗篇,如《赤壁怀古》中的“豪情顿减乘绿船,欲寄明主李后尘”等。
3. 陆游: 陆游是南宋时期的文学家、茶文化爱好者。他以喝茶、写诗为乐,他的《茶坊记》记述了他品茗的心得和对茶的热爱。
4. 王安石: 王安石是北宋时期的政治家、文学家,他在茶文化发展方面做出了贡献。他积极推广茶种植技术,提倡茶叶的广泛普及。
5. 宋徽宗: 宋徽宗是北宋时期的皇帝,他对茶文化的热爱也是众所周知的。他亲自制作红茶,并将其命名为“神品”。

在宋代,不同阶层的人们都有饮茶的习惯。士人、文人、官员和贵族常常聚集在茶馆或茶坊,品茶交流文化。而普通百姓则将茶作为日常生活的一部分。

此外,宋代还涌现了许多与茶相关的作品,如《茶经》、《茶论》等。这些作品通过介绍茶的分类、质量评定、煮茶技巧和茶具的使用等内容,进一步丰富了茶文化的内涵。总的来说,宋代是茶文化盛行的时期,有许多与茶相关的人物和作品,他们的贡献推动了茶文化的发展和传承。

图 8 直接询问 ChatGPT3.5 的结果
Figure 8 Direct result of ChatGPT3.5

宋代是中国茶文化的最高峰,在当时的社会中无论贵贱,饮茶已经成为日常生活常态,并且形成了以茶艺术为代表的优雅的茶文化。当时的知名茶文化人物有诸如诗人苏轼、陆游、茶学家曾公亮等。苏轼的《赋得古原草送别》中“山水浓淡皆殊质,茶亦淡中有浓香”、陆游的《示儿》中“论诗非宋昌伯,品茶只唤陆仲元”等诗句表现出他们对于茶的独到见解与热爱。

宋代饮茶的阶级差异也蛮明显,贵族和士人通常会在雅致的环境中,通过煮沸、烹炒、烘焙、研磨等复杂的工序来饮茶,对茶叶的选择也更为挑剔。而平民百姓则更多地选择泡茶,茶饮的环境和氛围也相对较为简单。

此外,宋代和茶有关的作品除了上述的诗歌外,还有《大观茶论》、《茶录》等重要的茶书。这些作品非常详尽的介绍了宋代的茶文化,无论是茶叶的种植、制作工艺,还是茶的品种,饮茶的道具和方式等内容。

总体来说,宋代可以说是中国茶文化的“龙凤盛世”。

图 9 直接询问 ChatGPT4 的结果
Figure 9 Direct result of ChatGPT4

宋代是中国历史上茶文化发展的一个重要时期,称为“龙凤盛世”。这一时期茶文化盛行,各个阶层都参与其中,并且产生了许多与茶相关的文学作品、艺术和社会活动。以下是与宋代茶文化相关的人物、阶层饮茶情况及著名的茶文化作品:

1. 茶文化相关人物

宋代有许多与茶文化相关的知名人物,根据人名库提供的数据,我们可以找到一些代表性人物:

- **蔡襄**: 蔡襄是宋代的茶学家和书法家,他撰写了《茶录》,对茶的生产、鉴别、冲泡,以及茶具等方面进行了详细的描述。
- **苏轼**: 苏轼不仅是著名的诗人、词人,也是嗜茶之人。他的许多诗词和散文中都涉及茶文化,如《试院煎茶》、《汲江煎茶》等。
- **陆游**: 陆游也是一位茶文化爱好者,他的诗中多次提到茶,如《夜坐试茶》、《尝新茶》等。

2. 宋代不同阶层的饮茶风貌

- **皇室贵族**: 宋代皇室对茶文化极为重视,甚至有专门的煎茶法和茶道仪式。例如宋徽宗赵佶在他的《大观茶论》中对茶的鉴别和品评作了详细论述。
- **文人雅士**: 文人雅士是茶文化的主要推动者和受益者。他们通过诗词、书画等方式表达对茶的热爱,并将饮茶提升为一种精神享受。
- **士农工商**: 普通百姓也参与到饮茶行列中,茶作为日常生活中的饮品,已广泛普及。宋代茶馆遍布城市和乡村,成为人们社交的重要场所。

3. 与茶有关的作品

宋代不仅是文学的黄金时代,也是茶文化作品的繁荣期。以下是一些著名的茶文化作品:

- **《大观茶论》**: 作者宋徽宗赵佶,是一部详尽的茶学理论书籍,对茶树的品种、采摘、制造等方面进行了系统论述。
- **《茶录》**: 蔡襄所著,是一部专门论述茶的生产、鉴别、冲泡及茶具的著作。
- **《东坡七体茶诗》**: 苏轼的诗集,其中多篇涉及茶的酬酢和品赏,是茶文化与文学结合的经典之作。
- **《西园雅集图》**: 由宋代画家李公麟绘制,展现了文人雅士雅集、品茶的场景。

总结起来,宋代无论从人物、社会阶层还是文化作品上,都呈现出一种茶文化的繁荣之态,被称作“龙凤盛世”实至名归。

图 10 直接询问 ChatGPT4o 的结果
Figure 10 Direct result of ChatGPT4o

这可能与模型自身能力有关系。另外,在长上下文的理解方面,ChatGPT3.5 和 ChatGPT4 存在一定的错误,可能是模型在 RAG 的过程中直接使用了 Google 检索结果的文本导致的。

从实验 2 来看,相较于 RAT,让模型直接回答的问题,仅完成了对模型自身的知识进行回答,无法结合本地数据进行。

从实验 1 和实验 2 的实验过程观察以及实验结

果来看, RAT 在知识挖掘服务的应用方面依赖于大语言模型的推理能力和检索效果。对于较弱的语言模型, RAT 的提升效果具有一定的局限性。另外, RAT 在知识挖掘方面依赖外部知识库的质量和覆盖范围, 如果知识库存在局限性, 则直接会影响 RAT 的检索效果和生成结果。另外, 由于 RAT 采用了明确的步骤式推理, 搜索出所有的内容进行结果生成, 步骤越长, 对于模型的输入要求就越大, 但由于模型的输入并不是无限的, 其生成结果受模型的输入长度的限制, 且由于输入越大模型性能下降越大, 因此需要控制 RAT 最终生成结果的输入。通过本实验结果来看, 虽然 RAT 在知识挖掘方面具有一定的局限性, 但也证明了 RAT 在大语言模型应用于知识挖掘服务的可行性。

4.4 实验局限性及改进方向

本实验旨在探索基于检索增强思维的图书馆知识挖掘服务, 通过融合 CoT 与 RAG 的 RAT 方法, 展现了基于检索增强思维的大语言模型在图书馆知识挖掘服务中的应用潜力。通过对实验结果的深入分析, 研究揭示了该系统的潜在优势, 也同时发现了一些局限性, 部分问题在实验中进行了改进尝试, 以下从 4 个方面进行详细分析。

①对于基于零样本 (zero-shot) 的任务评估问题, 在实验的草稿部分, 采用了零样本评估方法, 即仅依据提示词引导大型语言模型生成对应的草稿。这种方式存在一定的挖掘方向不确定性, 影响了挖掘结果的可控性。为解决这一问题, 可以在提示词中加入更具体的挖掘方向指引, 以规范挖掘方向。此外, 通过让大语言模型阅读相关引用文献, 约束模型挖掘方向, 也是一个可行的方法。

②对于知识库的依赖性问题, 实验发现 RAT 的表现与外部知识库的质量和覆盖范围密切相关。当知识库存在局限性时, 会对 RAT 生成结果产生影响。例如, 在实验中错误引用了人名规范库中关于“宋朝”和“白茶”两个人物的信息。因此为确保生成结果的准确性和可信度, 可以在每段稿件完成后引入反思流程, 让大语言模型检查并评估撰写内容, 对撰写的内容进行检查和校对, 确保其逻辑性和学术性。

③对于检索策略的影响问题, 每段稿件撰写依赖于从草稿中提取的中心思想和关键词。如果未能检索到相关信息, 大型语言模型将自行扩展和续写, 在这种情况下生成结果依赖于模型自身的能力 (这个过程中自然有可能继承了大语言模型的幻觉, 影

响生成内容的准确性)。为避免出现无法检索到内容的情况, 可以建立一个独立的检索系统, 为大型语言模型提供更精确的检索服务, 通过扩展检索关键词和优化检索结果排序来提高命中率。同时, 构建一个适应系统实际情况的知识图谱, 有助于提升检索准确性和有效性。

④对于模型输入长度的局限性问题, 由于模型的输入长度有限, 当撰写过程中涉及较多参考文献, 或者前后文内容较长时, 会导致模型在上下文推理中出现一定的“幻觉”, 即生成非真实或不一致的内容。对于上下文内容过长的问题, 可以通过让大语言模型对之前的对话内容进行总结和提炼来克服, 尽管这可能会导致一定的信息丢失。对于参考文献存在冲突的情况, 只能通过提升文献的质量来解决, 从源头上确保参考资料的准确性和一致性。

5 检索增强思维在图书馆知识挖掘中的价值 / Value of RAT in library knowledge mining services

5.1 检索增强思维在图书馆知识服务的应用场景

检索增强思维在复杂推理与生成中的独特优势, 在复杂资源分类、参考咨询服务和文献整理与数字人文服务方面提供了新的应用场景。

在复杂资源分类中的价值方面, 由于在跨学科资源的分类过程中, 传统方法由于受限于固定的分类体系, 面临资源分类体系多样化、资源揭示不充分、交叉学科资源分散以及难以揭示新学科资源的等问题^[26]。RAT 通过其多步推理能力, 可在分析资源内容的同时, 梳理其跨学科关联, 为资源的精确分类提供支持。例如, RAT 可以动态生成标签式的分类结果, 标注资源在不同领域的多重属性, 从而更全面地反映其特性。RAT 也可以根据资源内容自动分析和构建资源集。

在参考咨询服务中的优势方面, 由于复杂的参考咨询需求通常涉及多层次信息的分析与整合。RAT 通过迭代修订生成的思维链, 可以模拟馆员的思考过程, 从而为读者提供更准确和全面的答案^[8]。例如, 在解答涉及历史事件或技术演变的问题时, RAT 可以以馆员的思考过程来综合多种信息源, 为用户呈现详尽的背景信息及相关资料。

在二次文献的整理与加工方面, 馆员或者研究人员, 在分析和检索相关文献资料时, 常常需要处理大量的原始文献, 并将其根据不同的研究主题整理成系

统化的二次文献库。RAT 在这一过程中能够处理需要多步推理的复杂任务,辅助研究人员高效地分析文献,筛选出有价值的资料,并将其整合为结构化的二次文献库或是形成主题策展、展览、图书馆读者活动等。这不仅提高了文献整理的效率,还促进了知识的有效利用,为后续的研究提供了坚实的基础。

5.2 检索增强思维在图书馆知识挖掘中的价值

图书馆的知识服务体系中,检索增强思维是优化知识挖掘流程、提升服务效能的重要驱动力,在优化知识组织、促进知识发现及增强知识图谱构建等方面具有一定的应用价值。

在优化知识组织与检索服务方面,在优化知识组织与检索服务方面,传统的知识组织方式主要依赖预定义的分类体系,这不仅无法展示人类创造知识的动态逻辑过程,也无法根据知识创造的特点提供相应的知识和信息,难以满足现代图书馆资源在多样性与动态性上的需求。采用可解释推理方法的 RAT,凭借其智能推理和数据分析能力,能够动态识别数据中的模式与关联,构建类似线索的知识网络。这意味着图书馆不再仅依赖预先设定的目录结构,而是能够根据用户需求和资源特性,提供更具适应性和关联性的知识组织方案。同时,通过可解释的推理过程,用户可以清晰理解知识组织和检索的逻辑,提高系统的透明度和可信度,从而更有效地满足用户的多样化信息需求^[27]。

在增强知识图谱的构建与分析能力方面,RAT 有助于知识图谱数据的利用和关系分析。现代图书馆逐渐转向知识图谱的构建,通过数据之间的关联来展示更为立体化的知识结构^[28-29]。RAT 结合知识图不仅能够有效构建这些复杂的关系网络,还能通过多步推理帮助图书馆员和用户理解数据之间更深层次的关联。例如,从不同类型的资源中推理出某一历史事件的时间线、相关人物、影响范围等,从而为读者提供全面的信息支持,为可视化展示提供数据支撑。

6 结论与展望 /Conclusion

本文探讨了大语言模型时代图书馆知识挖掘服务的发展趋势,重点分析了检索增强思维技术在提升知识发现中的应用。通过上海图书馆人名规范库的宋代茶文化相关数据进行挖掘,不仅展示了知识挖掘的技术路径,也为智慧图书馆建设提供了实践案例。实验结果表明,其有助于提升知识挖掘服务的效率,

但其效果依赖于大语言模型的推理能力及知识库的丰富性和准确性。通过人工评价,检索增强思维知识挖掘方面具有一定的可行性。

具体而言,检索增强思维能够高效挖掘相关历史人物和事件,整合多源资料,深入分析数据逻辑,科学解读文化现象,并提供创新研究视角。然而,检索增强思维在处理复杂推理和依赖知识库方面仍存在局限。未来研究应优化检索增强思维技术以应对更复杂的知识挖掘任务,提升模型的适应性及泛化能力,并结合其他人工智能技术,实现更智能化、个性化的知识服务。

基于检索增强思维的知识挖掘研究为图书馆知识服务和智慧图书馆的转型升级提供了新的技术方案。图书馆在数字化和 AIGC 时代需持续创新,探索多样化的知识服务模式,以应对知识爆炸和信息环境的变化,进一步提升其在知识管理与服务中的核心作用。

参考文献 /References:

- [1] PAN Y. Heading toward artificial intelligence 2.0[J]. Engineering, 2016, 2(4): 409-413.
- [2] 张晓林. 颠覆性变革与后图书馆时代——推动知识服务的供给侧结构性改革[J]. 中国图书馆学报, 2018, 44(1): 4-16. (ZHANG X L. Disruptive changes and the post-library era: toward supply-side structure reform of knowledge services[J]. Journal of library science in China, 2018, 44(1): 4-16.)
- [3] 左静远. 基于知识挖掘技术的公共图书馆智慧服务调查与思考[J]. 情报探索, 2023(9): 100-107. (ZUO J Y. Investigation and reflection on intelligent service of public library based on knowledge mining technology [J]. Information research, 2023(9): 100-107.)
- [4] 郭利敏, 付雅明. 融合 ReAct 模式的图书馆大语言模型知识服务系统构建[J]. 图书馆论坛, 2024, 44(6): 61-70. (GUO L M, FU Y M. An LLM-based knowledge service system for libraries integrating the ReAct model [J]. Library tribune, 2024, 44(6): 61-70.)
- [5] ZHUANG Y, WU F, CHEN C, et al. Challenges and opportunities: from big data to knowledge in AI 2.0[J]. Frontiers of information technology & electronic engineering, 2017, 18: 3-14.
- [6] 陈帅朴, 刘芳霖, 钱宇星, 等. 检入新境: 大语言模型引领的信息检索主题与知识关联演化分析[J/OL]. 图书情报知识, 1-14[2024-12-16]. (CHEN S P, LIU F L, QIAN Y X, et al. Topic and knowledge association evolution in the field of large language model-enabled information retrieval[J/OL]. Documentation, information & knowledge, 1-14[2024-12-16].)

- [7] 张宏玲, 沈立力, 韩春磊, 等. 大语言模型对图书馆数字人文工作的挑战及应对思考[J]. 图书馆杂志, 2023, 42(11): 31-39, 61. (HONG L Z, SHEN L L, HAN C L, et al. Challenges and reflections on the practical application of large language model in digital humanities work at libraries [J]. Library journal, 2023, 42(11): 31-39, 61.)
- [8] WANG Z, LIU A, LIN H, et al. Rat: retrieval augmented thoughts elicit context-aware reasoning in long-horizon generation[EB/OL]. [2025-03-17]. <https://arxiv.org/abs/2403.05313>.
- [9] 王颖纯, 董雪敏, 刘燕权. 基于知识挖掘的图书馆智慧推荐服务模式[J]. 图书馆学研究, 2018(9): 37-43. (WANG C Y, DONG X M, LIU Y Q. Service mode of smart recommendation based on knowledge mining[J]. Research on library science, 2018(9): 37-43.)
- [10] 张晓林. 重新认识知识过程 and 知识服务[J]. 图书情报工作, 2009, 53(1): 6-8. (ZHANG X L. Reconceiving the knowledge process and knowledge service[J]. Library and information service, 2009, 53(1): 6-8.)
- [11] 温有奎, 温浩, 乔晓东. 让知识产生智慧——基于人工智能的文本挖掘与问答技术研究[J]. 情报学报, 2019, 38(7): 722-730. (WEN Y K, WEN H, QIAO X D. Research on the methods of information Science and artificial intelligence fusion innovation[J]. Journal of the China society for scientific and technical information, 2019, 38(7): 722-730.)
- [12] 罗威, 谭玉珊. 基于内容的科技文献大数据挖掘与应用[J]. 情报理论与实践, 2021, 44(6): 154-157. (LUO W, TAN Y S. Content-based data mining and application of scientific and technical literature big data[J]. Information studies: theory & application, 2021, 44(6): 154-157.)
- [13] 王露莹. 大数据时代图书馆馆藏资源的知识标注与知识挖掘研究[J]. 农业图书情报学刊, 2018, 30(10): 70-73. (WANG L Y. Research on knowledge annotation and mining of library collection resource in the era of big data[J]. Journal of library and information science in agriculture, 2018, 30(10): 70-73.)
- [14] 胡昊天, 邓三鸿, 孔玲, 等. 生成式情报学术语自动抽取与多维关联知识挖掘研究[J]. 情报学报, 2024, 43(5): 588-600. (WANG L Y, DENG S H, KONG L, et al. Automatic generative information science term extraction and multidimensional linked knowledge mining[J]. Journal of the China society for scientific and technical information, 2024, 43(5): 588-600.)
- [15] BERTI A, QAFARI M S. Leveraging large language models (llms) for process mining (technical report)[EB/OL]. [2025-03-17]. <https://arxiv.org/abs/2307.12701>.
- [16] JOSÉ C D J, RACHEL H, RICHARD S, et al. Domain-driven LLM development: insights into RAG and fine-tuning practices[C]// In proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining (KDD' 24). New York: Association for computing machinery, 2024: 6416-6417.
- [17] WAN M, SAFAVI T, JAUHAR S K, et al. TNT-LLM: text mining at scale with large language models[C]//Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining. New York: Association for Computing Machinery, 2024: 5836-5847.
- [18] UZUNOGLU A, SAFA A R, ŞAHİN G G. Paradise: evaluating implicit planning skills of language models with procedural warnings and tips dataset[EB/OL]. [2025-03-17]. <https://arxiv.org/abs/2403.03167>.
- [19] ZHANG Z, FANG M, CHEN L, et al. How do large language models capture the ever-changing world knowledge? a review of recent advances[EB/OL]. [2025-03-17]. <https://arxiv.org/abs/2310.07343>.
- [20] 黄永文, 孙坦, 赵瑞雪, 等. 大数据与人工智能背景下新型知识服务研究与实践[J]. 图书情报工作, 2022, 66(19): 36-46. (HUANG Y W, SUN T, ZHAO R X, et al. Research and practice of new knowledge service in the context of big data and artificial intelligence[J]. Library and information service, 2022, 66(19): 36-46.)
- [21] 赵月, 何锦雯, 朱申辰, 等. 大语言模型安全现状与挑战[J]. 计算机科学, 2024, 51(1): 68-71. (ZHAO Y, HE J W, ZHU S C, et al. Security of large language models: current status and challenges [J]. Library and information service, 2022, 66(19): 36-46.)
- [22] LEWIS P, PEREZ E, PIKTUS A, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks[J]. Advances in neural information processing systems, 2020, 33: 9459-9474.
- [23] YU H, GAN A, ZHANG K, et al. Evaluation of retrieval-augmented generation: a survey[C]//CCF conference on big data. Singapore: springer nature Singapore, 2024: 102-120.
- [24] WEI J, WANG X, SCHUURMANS D, et al. Chain-of-thought prompting elicits reasoning in large language models[J]. Advances in neural information processing systems, 2022, 35: 24824-24837.
- [25] CHU Z, CHEN J, CHEN Q, et al. Navigate through enigmatic labyrinth a survey of chain of thought reasoning: advances, frontiers and future[EB/OL]. [2025-03-17]. <https://arxiv.org/abs/2309.15402>.
- [26] 邓盼盼, 李军莲, 冀玉静, 等. NSTL 科技文献资源分类体系构建研究[J]. 数字图书馆论坛, 2024, 20(8): 1-8. (DENG P P, LI J L, JI Y J, et al. Construction of NSTL scientific and technological literature resource classification system[J]. Digital library forum, 2024, 20(8): 1-8.)

- [27] 夏毅, 兰明敬, 陈晓慧, 等. 可解释的知识图谱推理方法综述[J]. 网络与信息安全学报, 2022, 8(5): 1-25. (XIA Y, LAN M J, CHEN X H, et al. Survey on explainable knowledge graph reasoning methods[J]. Chinese journal of network and information security, 2022, 8(5): 1-25.)
- [28] 王娟, 曹树金, 王志红, 等. 面向探索式搜索的领域知识图谱构建及实验探索[J]. 图书情报工作, 2024, 68(3): 105-116. (WANG J, CAO S J, WANG Z H, et al. Construction and experimental exploration of domain knowledge graph for exploratory search[J]. Library and information service, 2024,

68(3): 105-116.)

- [29] 田杰, 周军兰. 智慧图书馆模型与系统体系构建探析[J]. 文献与数据学报, 2020, 2(2): 40-48. (TIAN J, ZHOU J L. A study on the construction of smart library model and system[J]. Journal of library and data, 2020, 2(2): 40-48.)

作者贡献说明 / Author contributions:

郭利敏: 提出研究思路、实验设计、论文写作及修改

付雅明: 实验设计、论文写作及修改

刘悦如: 实验结果核验、论文修改

Research on the Application of Retrieval Augmented Thoughts in Library Knowledge Mining Services*

Guo Limin¹ Fu Yaming² Liu Yueru³

¹Shanghai Library/Institute of Scientific & Technical Information of Shanghai, Shanghai 200031

²School of Cultural Heritage and Information Management, Shanghai University, Shanghai 200444

³Tongji University Library, Shanghai 200092

Abstract: [Purpose/Significance] The study focuses on the development of library knowledge mining services in the era of large language models. It explores the integration of traditional retrieval methods with retrieval augmented thoughts (RAT) to provide users with both profound and logically rigorous knowledge information, aiming to enhance the quality of library knowledge services. [Method/Process] The study adopted an experimental research approach, taking the Name Authority Database of Shanghai Library as the data source. It specifically focused on mining information related to tea culture in Song Dynasty, uncovering more details about this historical period from library collections, thus offering users a more comprehensive and in-depth perspective. [Result/Conclusion] Experimental results indicate that the RAT method has certain advantages in library knowledge mining services. This approach not only improves the efficiency and accuracy of data mining but also provides new technical solutions for library knowledge mining, offering valuable insights for the development of smart libraries.

Keywords: knowledge mining retrieval augmented thoughts (RAT) smart library services large language models

*This work is supported by Shanghai Magnolia Talent Program Pujiang Project titled "Research on Cultural Heritage Knowledge Service Based on Generative Artificial Intelligence"(Grant No. 24PJC036) and the Philosophy and Social Sciences of Shanghai project titled "Smart Data-Driven Research Innovation Support Services in Metropolitan Libraries"(Grant No. 2020BTQ005).

Author(s): Guo Limin, associate research librarian, master; Fu Yaming, assistant researcher, PhD, corresponding author, E-mail: ymingfu@126.com; Liu Yueru, associate research librarian, master.

Received: 2024-10-15 Revised: 2024-12-17