文献衍生的聚合物知识的检索增强生成：以可生物降解聚合物专家系统为例

Sonakshi Gupta，[1] Akhlak Mahmood，[2] Wei Xiong，[3] 和 Rampi Ramprasad[3]，*[1]佐治亚理工学院计算科学与工程学院，美国佐治亚州亚特兰大市，邮编*30332，771 FerstDriveNW*。[2]*Matmerize*公司，美国佐治亚州亚特兰大市，邮编*30308*。[3]佐治亚理工学院材料科学与工程学院，美国佐治亚州亚特兰大市，邮编*30332，771 Ferst Drive NW*

arXiv:2602.16650v1 [cs.CE] 18 Feb 2026

摘要

聚合物文献包含大量且不断增长的实验知识，但其中许多内容埋藏在非结构化文本中，并且采用不一致的术语进行报告，因此难以系统地检索、比较或推理。现有工具通常只提取范围狭窄、针对特定研究的事实，却未能提供回答跨多项研究的科学问题所需的综合上下文。检索增强生成（RAG）通过将大语言模型（LLM）与外部检索相结合，为克服这一局限提供了有前景的途径，但其有效性很大程度上取决于领域知识的表征方式。在本研究中，我们为材料科学语料库开发并调整了两种检索流程：一种是基于密集语义向量的方法（VectorRAG），另一种是基于图的方法（GraphRAG）。我们利用 1,000 多篇聚羟基烷酸酯（PHA）论文，构建了保留上下文的段落嵌入，以及经过规范化的结构化知识图谱，以支持实体消歧和多跳推理。为评估这些流程，我们使用检索指标开展了广泛的基准测试，将其与 GPT 和 Gemini 等通用先进系统进行比较，并由一位领域化学家进行定性验证。结果表明，GraphRAG 的精确率和可解释性更高，而 VectorRAG 的召回率更广，凸显了两种方法之间互补的权衡。专家验证进一步证实，这些定制流程，尤其是 GraphRAG，能够生成依据充分、引文可靠且与领域高度相关的回答。通过让每项陈述都以明确的来源证据为依据，这些系统为研究人员提供了一种可靠的方式，用于查阅文献、比较不同研究的发现，并揭示难以人工提取的规律。更广泛而言，本研究展示了一条切实可行的路径：利用经过整理的语料库和检索设计，开发材料科学领域的学术助手，从而减少对大型专有模型的依赖，并实现可普及、可信赖且可规模化的文献分析。

MAIN

数据长期以来一直被视为材料信息学的基石 [1——3], 随着该领域近期进展不断依赖数据驱动技术，这一点也愈发明显。从算法设计到自主发现工作流，近期的突破

* rampi.ramprasad@mse.gatech.edu

不仅得益于机器学习的改进，也得益于大型、高质量材料数据集的可用性。然而，对于许多聚合物体系而言，此类数据集仍然有限或不可获取。相关信息主要见于浩瀚、异质且大多未经结构化的文献之中；这些文献仍以前所未有的速度不断扩充，数十年来积累的科学知识散见于文本、图表和表格中 [4]。因此，这些信息中的大部分难以检索、整合不同研究的成果，或有效用于科学推理。

以往的工作试图挖掘这一资源库 [5–9], ，从而生成可支持下游任务的结构化数据，例如性能预测 [10, 11], 、合成规划 [12, 13], 以及材料设计 [14, 15]。尽管这些数据库提供了所提取知识的宝贵静态快照，但它们并未充分解决聚合物科学家在实践中如何与文献互动的问题。科学探究往往需要跨研究动态检索证据、调和不一致的术语，并生成在不同实验背景下报告的结果。因此，即便是看似直截了当的问题，例如“在不同加工条件下，已报道的聚羟基烷酸酯薄膜气体渗透性数值有哪些？”或“在酸性条件与碱性条件下，PHB水解会形成哪些产物？”，通常也无法通过查阅单篇论文或单条数据库记录来回答。相反，回答此类查询需要识别相关实验背景，并整合多个来源中报告的碎片化发现，这凸显了对支持基于文献的推理而非孤立事实提取的系统的需求。

因此，一位高效的聚合物文献研究者必须超越信息提取，做到：（i）从多篇论文中检索相关证据；（ii）在综合信息的同时保留实验背景；（iii）当文献未提供充分证据时，适时暂缓判断；（iv）明确以可核实的原始资料为结论提供依据，以便领域专家评估其可靠性和可信度。

大语言模型（LLM）能够利用从海量文本语料库中学到的知识，强力整合分散的科学信息。然而，大语言模型一个有据可查的局限是幻觉：它们会生成流畅、自信但事实错误的陈述 [16, 17]。近期研究 [18] 表明，幻觉之所以出现，是因为大语言模型的优化目标是续写文本，而非验证信息；标准训练目标甚至会奖励自信作答，即便相关证据不足——

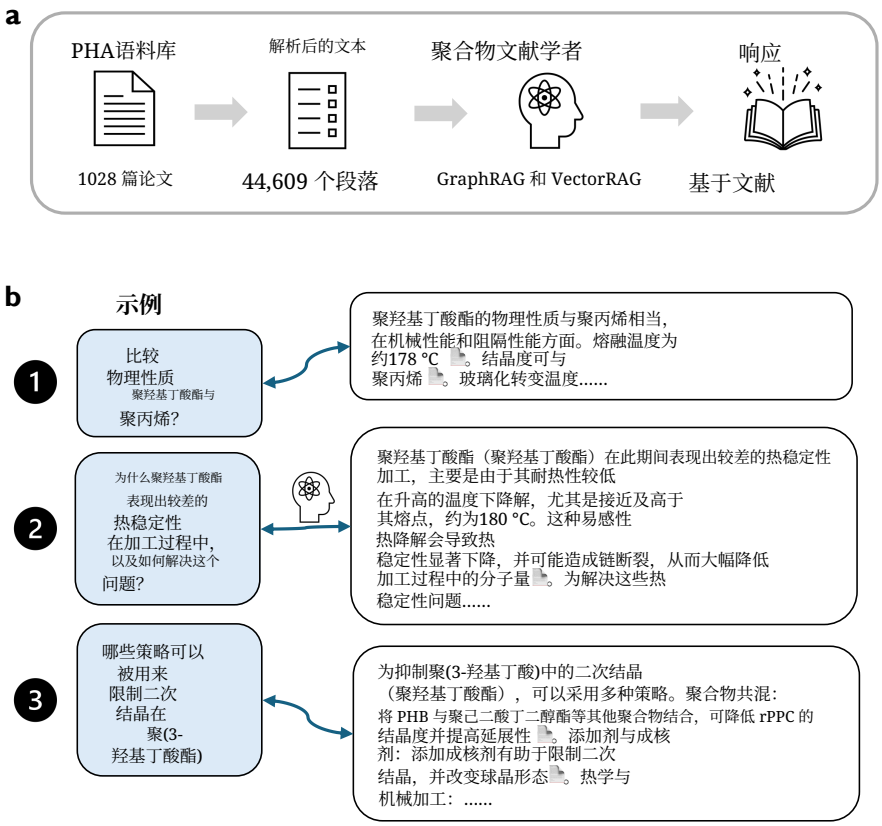

图 1. (a) 面向聚羟基脂肪酸酯文献依据型问答的检索增强生成（RAG）框架概览。经过整理的 PHA 语料库由 1,028 篇文章和 44,609 个解析后的段落组成，是 VectorRAG 和 GraphRAG 检索流程共用的知识库。每种流程检索到的上下文证据都会提供给大语言模型，以生成基于文献依据的回答。(b) 聚合物文献学者生成的科学查询示例及相应回答。这些回答通过汇总从 PHA 语料库中的多项相关研究检索到的证据合成，如文档符号所示。

支持证据薄弱。这种行为在科学场景中尤为令人担忧，因为如果没有明确的依据支撑，错误陈述可能很难被发现。为减少幻觉并提高可靠性，研究人员探索了多种策略，包括提示工程、在生成过程中优化模型输出、自反馈机制，以及检索增强生成（RAG）方法 [19–22]。在检索增强生成系

统、模型从可信的外部来源检索相关信息，并依据这些证据生成回答，从而减少缺乏依据的说法，同时提升透明度和可追溯性；因此，检索增强生成特别适用于科学领域。

近年来，检索增强生成方法发展迅速，这得益于多项进展：检索与生成的结合更加紧密，模型也能更有效地利用外部知识 [23]。这些进展催生了两种突出的检索范式：基于向量的稠密检索和基于图的推理 [24]。基于向量的检索增强生成（Vector- RAG）将信息表示为高维空间中的稠密嵌入，并通过相似性搜索检索语义相似的段落；大量研究致力于优化分块、嵌入模型、索引和评分机制 [25]。相比之下，基于图的检索增强生成（GraphRAG）将知识组织为由相互关联的实体和关系构成的网络，从而能够通过图遍历或子图提取进行检索。这种结构化表示支持多跳推理，提升了关系依赖复杂领域中的检索精确率，并生成更易于理解的证据链 [26]。这两种互补的检索范式共同奠定了基础，有助于支持材料科学领域中可靠且有据可依的大语言模型推理；在该领域，准确检索证据和进行可解释的关系推理都至关重要。

将这些检索策略应用于材料科学领域，会带来独特的挑战。该领域的科学文本高度异质，充斥着复杂的专业术语、层层嵌套的性质—组成—加工关系，以及命名法不一致等问题。例如，同一种聚合物可能被称为“PLA”“聚乳酸”或“聚丙交酯”，这使得汇总和比较不同研究报告的结果更加困难。对于聚羟基脂肪酸酯（PHAs）而言，这些挑战尤为突出。PHAs 是种类多样、发展迅速的一类可生物降解聚合物，在可持续材料设计方面备受关注 [27]。因此，PHAs 为评估 VectorRAG 和 GraphRAG 等互补检索策略提供了一个极具代表性的测试案例。

在本研究中，基于图 1 所示的框架，我们推出了聚合物文献学者，这是一个检索增强系统，旨在支持高分子科学中基于文献的推理。我们在一个包含 1,000 多篇聚羟基脂肪酸酯全文论文的精选语料库上，针对该领域定制了 VectorRAG 和 GraphRAG 流程，并对其进行了系统基准测试。VectorRAG 通过领域感知的分块和嵌入策略进行调整，GraphRAG 则通过稳健的实体抽取、规范化以及

规范化流程进行调整，以支持可扩展的图推理。为系统评估检索质量和事实依据支撑，我们整理了一个包含 113 个由领域专家评估的问题的基准测试，问题聚焦于聚羟基脂肪酸酯，并进行了定性专家分析，涵盖通用科学查询、论文特定问题和多论文推理任务。我们依据内容准确性、引文完整性和科学推理深度，对五种代表性系统生成的回答进行了评估：搭配 GPT-4o-mini 的 GraphRAG、搭配 Llama-3.1-70B 的 GraphRAG、搭配 GPT-4o-mini 的 VectorRAG、采用 ChatGPT-5（网页搜索）的 OpenAI 内置检索增强生成，以及采用 Gemini 的 Google 内置检索增强生成。

为了支持系统的实际应用和专家评估，我们还开发了一个交互式界面，用于展示 VectorRAG 和 GraphRAG 检索流程及其底层证据。领域专家可以通过该界面查询系统、查看检索段落或知识图谱三元组、核实证据来源，并评估生成回复的可信度和完整性。

本研究的一项重要贡献，是展示了如何调整最先进的检索方法，仅凭经过精心整理的语料库和适度的计算资源，构建可靠、专注于特定领域的聚合物文献学者，而无需依赖通用基础模型通常所需的规模或成本。通过提出透明且可复现的检索增强框架，我们表明，可靠地推理聚合物文献并不需要大型专有模型。即使使用规模较小的开放权重模型，该系统依然有效，能够明确以文献为依据作答，并在现有证据不足时恰当地回答“我不知道”。我们以聚羟基脂肪酸酯为例，展示了如何使基于向量和基于图的检索与高分子科学文献的结构相契合。VectorRAG 具有广泛的语义覆盖范围，并能有效捕捉段落级上下文；GraphRAG 则支持围绕关联实体开展更有结构、具备关系感知能力的推理。这些方法结合起来，为针对复杂且报告方式不一致的聚合物数据开展推理提供了有依据且可解释的基础，使由文献驱动的聚合物研究兼具实用性与可信度。

结果与讨论

检索增强推理流程概述

图 1 概述了聚合物文献学者。这是一种检索增强框架和交互式应用，旨在支持对聚羟基脂肪酸酯实验文献进行科学推理。该系统不只是检索孤立的事实，还能综合生成连贯且以文献为依据的解释，反映多项研究中对聚合物性能、加工行为和结构–性能关系的讨论与阐释。

如图 1a 所示，该框架基于一个精选语料库，其中包含 1,028 篇聚焦于 PHA 化学、加工和性能的全文期刊文章。这些文章被解析为 44,609 个段落级文本片段，作为科学证据的基本单位。这种段落级表示保留了实验背景，而这对于材料科学家推理聚合物行为至关重要。为实现对这些文献的有效检索，聚合物文献学者支持两种互补的检索策略：VectorRAG 和 GraphRAG。

在 VectorRAG 处理流程中，段落级文本被嵌入连续的语义空间，因此即使不同论文使用不同术语描述相似概念，也能检索到与实验相关的段落。相比之下，GraphRAG 处理流程将提取出的实体（如聚合物体系、加工步骤和测得的性能）组织成结构化表示，明确捕捉文献中报告的关系。这种基于图的视图能够跨相互关联的陈述进行推理，从而整合多项研究中的证据，共同描述某种材料行为或趋势。

在推理过程中，用户通过已开发的应用程序界面向聚合物文献学者提交查询。系统通过两种检索路径之一获取相关证据，并将所得段落或子图作为上下文输入提供给大语言模型。随后，模型综合直接取自 PHA语料库中实验报告的观察结果，生成有依据的回答，使用户能够获得有文献支持的解释。

图 1b 展示了 展示通过该流程生成的代表性示例查询及回答

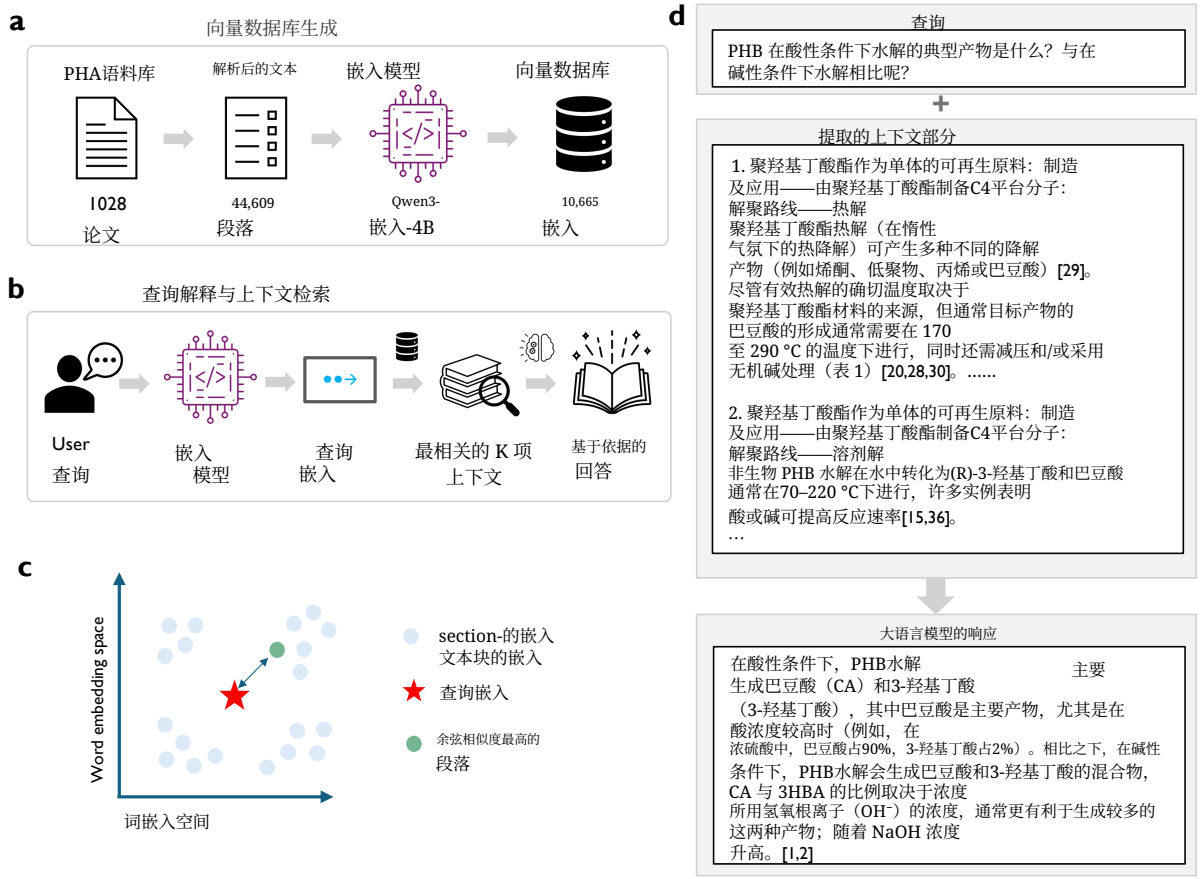

图2。 **用于基于文献的聚羟基脂肪酸酯问答的 VectorRAG 工作流程。**

(a) 语料库的后端处理：解析全文文章，按上下文归并为精简的文本块，并嵌入稠密语义空间，以便进行基于相似度的检索。**(b)** 检索增强推理：将用户查询编码到相同的潜在空间中，并与语料库中语义最匹配的文本片段进行匹配。**(c)** 嵌入空间的概念示意图，展示如何通过查询向量与其最近邻之间的余弦相似度识别相关段落。**(d)** 代表性查询示例，展示检索到的文献段落，以及语言模型生成的、基于依据且带引文链接的回答。

界面。每条回答并非引用单一来源，而是反映了多项相关研究的汇总证据，如图中的文档符号所示。这种汇总方式模拟了材料科学家通常查阅文献的方式：整合多篇论文中的观察结果，以形成对材

料行为的共识性理解，同时确保每条回答的证据依据透明可见。因此，该框架使科学家能够以符合实验实践的方式查询文献，从而获得基于文献的聚合物性能、加工考量和设计选择方面的见解；这些见解难以通过传统关键词搜索或数据库驱动的工具获得。

VectorRAG：稠密语义检索流程

VectorRAG 框架将文献语料库表示在一个高维语义空间中，该空间能够捕捉不同研究之间的上下文相似性。如图2所示，该工作流程整合了三个主要组成部分：知识库构建、基于语义相似性的检索，以及基于依据的答案生成。

在知识库构建阶段（图2a），44,609 个经过解析的段落通过合并同一节或小节中相邻的段落，组成保留上下文的文本块。这种方法确保描述相关概念（例如合成步骤和所得性质）的文本片段保持连贯，最终得到 10,665 个保留上下文的文本块。使用 *Qwen3-Embedding-4B* 模型将这些文本块嵌入为 3,584 维稠密向量，随后存储在关系型向量数据库中，以便高效地进行基于相似度的检索。

在查询处理期间（图2b–c），用户查询被嵌入到相同的向量空间中，并使用余弦相似度与语料库中的所有嵌入进行比较。检索到的文本片段数量（k）在 2 到 8 之间变化；当数量为 $k = 8$ 时，依据支撑效果和回答连贯性最佳。检索到的文本块与用户查询合并后提供给大语言模型，由其生成以文献为依据、并由显式引用支持的回答（图2d）。

这种密集语义检索方法能够有效捕捉隐含的概念关系和更广泛的叙事上下文，因此尤其适用于描述性或探索性问题，例如“哪些因素会影响 PHBV 共聚物的结晶度？”。然而，与此同时，检索到的文本缺乏结构，可能导致冗余，并在推理过程中增加词元使用量。因此，VectorRAG 提供了强有力的语义检索基线，可与 GraphRAG 的结构化关系推理能力互为补充。

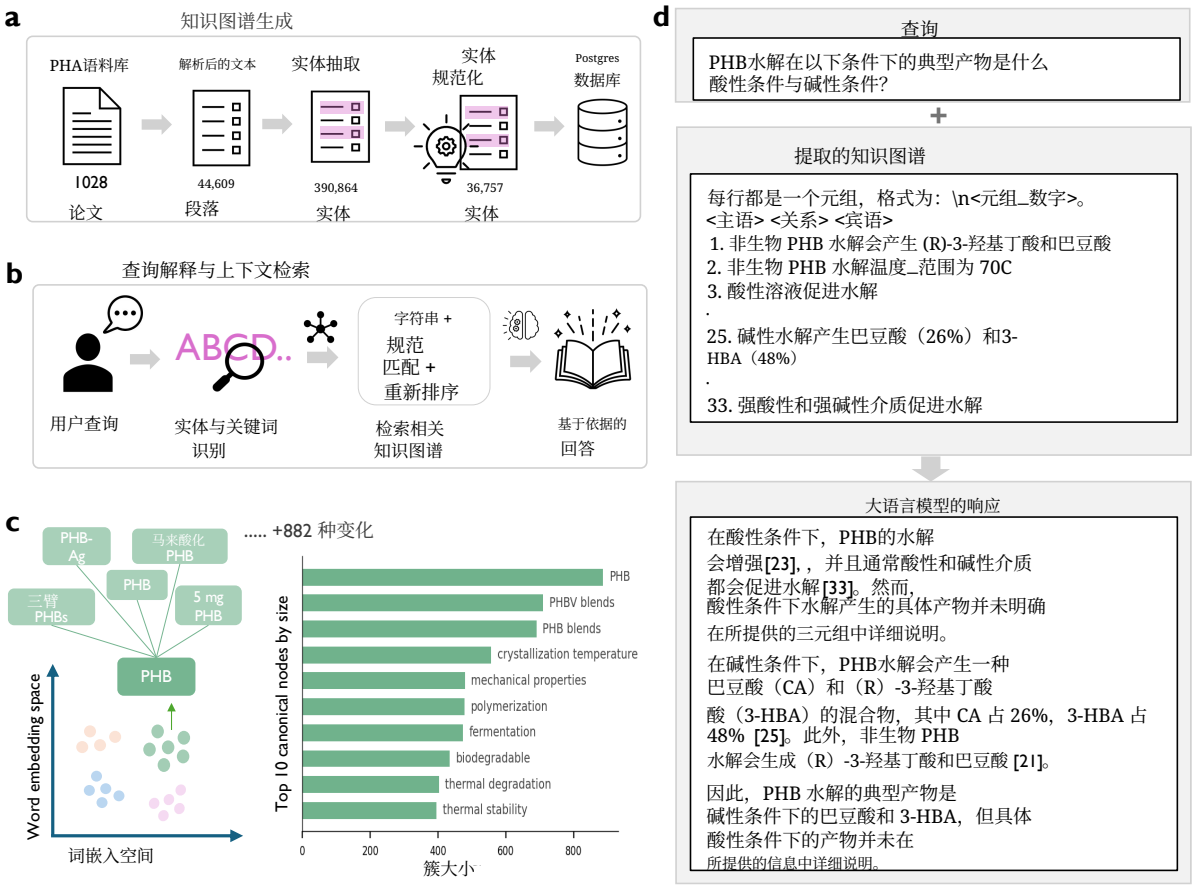

图 3. **面向聚羟基脂肪酸酯的基于知识图谱问答的 GraphRAG 工作流程。**

(a) 对语料库进行后端处理，从文献中提取实体和关系，对其进行规范化，并以关系元组的形式存储在关系数据库中。**(b)** 检索与推理阶段：将用户查询拆解为实体—关系对，与规范实体匹配，再通过基于路径的评分策略重新排序，以识别最相关的子图，用于有依据的响应生成。**(c)** 实体规范化示例：将多个相关提及（例如 PHB–Ag、马来酸酐化PHB、三臂PHB）合并为代表 PHB 的统一规范节点。配套的条形图突出显示了按簇大小排序的排名靠前的规范实体，展示了规范化如何提升图连通性和召回率。**(d)** 代表性示例，展示了用户查询、检索到的知识图谱三元组，以及语言模型综合支持性引文生成的相应基于依据的回答。

GraphRAG：基于知识图谱的检索流程

GraphRAG 框架将解析后的语料库重构为显式的知识图谱（KG），从而能够在科学文本中的多种关系之间进行可解释的多跳推理。如图 3 所示，该工作流程包括三个主要组成部分：知识图谱构建、查询解析与检索，以及基于依据的回答生成。

在后端构建阶段（图 3a），44,609 个解析后的段落经过实体抽取与规范化流程处理，具体如方法部分所述。研究评估了多个大语言模型，包括 GPT-4o、GPT-4o-mini、GPT-4.1-mini、Llama-3.1- 70B 和 Llama-3.3-70B，以评估模型选择如何影响元组质量和下游检索效果。其中，GPT-4o-mini 和 Llama-3.1-70B 生成的元组最为可靠，因此被选用于最终处理流程。该阶段使用的抽取提示及所测试的所有变体均见补充信息。

因此，GPT-4o-mini抽取了390,864个关系元组，Llama-3.1-70B抽取了311,475个元组。这些元组捕捉了诸如 `[PHB-hasproperty-Tg]`或 `[PHBV-synthesized with-hexanoate]`之类的科学关系。由于聚合物及相关实体常以多种文本变体出现，因此引入了实体规范化步骤，以合并等价提及并提高图连通性。该步骤用于合并等价提及，并使图更加互联。此步骤为GPT-4o-mini生成了36,757个规范实体，为Llama-3.1-70B生成了37,289个规范实体，从而提高了整个语料库的一致性与检索质量。所有所得元组均存储于关系数据库中，从而构建了源自聚羟基脂肪酸酯文献的结构化、可查询的科学知识表示。

在查询处理过程中（图 3b），系统使用实体与关键词解析器，将每个用户问题拆解为关键实体和关系。如方法部分所述，GraphRAG 结合规范实体匹配、字符串相似度和语义相似度来检索相关子图。随后，系统通过路径重排序策略对检索到的元组重新排序，并将子图大小上限设为 300 个节点。这些子图通常包含多条连接相关实体的路径，从而能够跨研究进行多跳推理。例如，对于“3HV 含量如何影响 PHBV 共聚物的熔融温度？”这样的查询，系统会从$PHBV \rightarrow 3HV \rightarrow Tm$开始遍历，检索所有相关

与这种关系相关的实验背景。

如图 3c 所示，该图将文本表述上的差异整合为规范节点，其中语境相关的实体（例如“三臂PHB”“马来酸化PHB”和“PHB–Ag”）均归并到单一规范实体 PHB 之下。图 3c 右侧面板展示了按簇大小排序的顶级规范节点，这些节点对应讨论频率最高的实体，例如 PHB、PHBV 共混物、结晶温度和降解行为。

最后，将检索到的子图上下文与用户查询拼接，并传递给大语言模型，以生成有引文依据的回答（图 3d）。通过将证据表示为明确的关系元组，GraphRAG 提供了透明的推理链，增强了可解释性和可追溯性。与 VectorRAG 相比，这种结构化检索范式更注重关系精确性和推理深度，因此尤其适用于需要跨研究进行多跳推理的机理类或比较类问题。

基准测试与评估

受控评估：检索保真度基准测试

作为初步测试，我们选取了 21 个与 PHA 相关的数字对象标识符，用于开发和优化两种检索流程。这一受控设置确保检索增强生成框架能够从聚焦的数据集中有效检索相关信息；该数据集中的每篇论文都与查询主题直接相关。在审阅这 21 个数字对象标识符所对应的全文后，我们手动编写了 113 个问题，以评估检索保真度，评估指标采用方法部分中所述的召回率和准确率。

这 21 篇全文文章被解析为 943 个段落，并经过嵌入后存储在向量数据库中，供 VectorRAG 流程使用。与此同时，实体抽取与规范化为 GraphRAG 流程生成了 10,814 个原始元组；使用 GPT-4o-mini 将其整合为 5,904 个规范实体节点，使用 Llama-3.1-70B 则将 8,744 个原始元组整合为 4,973 个规范节点。在扩展至完整语料库之前，这些数据集构成了用于测试检索精确率和依据质量的基准测试。

如表 I 所示，在这个聚焦于 21 个数字对象标识符的测试集上，两种检索流程的表现都非常出色，所有模型的召回率都接近完美。这些高召回率表明，当语料库范围明确且主题一致时，稠密

表 I. **受控评估**：报告的指标包括 GraphRAG 和 VectorRAG 在 21 个 PHA DOI 子集上的召回率、准确率，以及平均每次查询的响应时间和成本；两者分别采用 300 个元组和 8 个段落的优化后上下文限制。

模型	上下文限制	召回率	准确率	平均响应时间（秒）	平均总计成本（美元）
GraphRAG					
GPT-4o-mini	300 个元组	0.982	0.991	7.18	0.00097
Llama-3.1-70B	300 个元组	0.991	0.991	10.23	0.00000
VectorRAG					
GPT-4o-mini	8 个段落	0.973	1.000	16.43	0.00186
Llama-3.1-70B	8 个段落	0.973	1.000	41.75	0.00000

（VectorRAG）和结构化（GraphRAG）检索框架都能成功检索所有相关信息。得益于紧凑的关系表示，GraphRAG 检索速度更快，计算成本也更低；相比之下，VectorRAG 的延迟较高，主要是因为它处理的文本输入长得多，长度往往接近模型的 8,192 个词元上下文限制。这些结果表明，在受控条件下，两种流程都能有效地为模型响应提供依据，并为后续的大规模评估奠定可靠的基线。

规模化评估：语料库规模下的检索性能

为评估受控基准测试之外的可扩展性和泛化能力，我们在完整的 PHA语料库上评估了两种检索流程，该语料库包含 1,028 篇期刊文章全文。我们使用召回率和准确率衡量性能，并结合每次查询平均响应时间和推理总成本来评估效率，结果汇总于表 II。

扩展至整个语料库后，两种流程的召回率均有所下降，反映出随着搜索空间扩大，检索精确上下文的难度也随之增加。

VectorRAG 的降幅更为明显，召回率为 0.717，这表明随着嵌入数量增加、上下文重叠加剧，密集语义检索的区分能力会减弱。相比之下，GraphRAG 的召回率显著更高，使用 GPT-4o-mini 时达到 0.938；这得益于其结构化的实体关联检索，可将搜索范围限制在知识图谱中语义一致的区域。尽管存在这些差异，两种框架的总体答案准确率始终保持较高水平，约为 0.96–0.97，这表明即使未检索到预期段落，处理流程通常仍能生成基于文献的正确回答。这一观察凸显了检索精确率本身并不能完全决定下游答案质量。相较之下，GraphRAG 的关系表示在大规模场景下具有更强的鲁棒性和可解释性，彰显了结构化检索在大型异质科学语料库中的优势。

表 II. **大规模评估**：报告的指标包括召回率、准确率，以及 GraphRAG 和 VectorRAG 在 1028 个 PHA DOI 上的平均每次查询响应时间和成本，分别使用 300 个元组和 8 个段落的优化后的上下文限制。

模型	**上下文限制**	**召回率**	**准确率**	**平均响应时间（秒）**	**平均总成本（美元）**
GraphRAG					
GPT-4o-mini	300 个元组	0.938	0.973	34.14	0.00070
Llama-3.1-70B	300 个元组	0.903	0.964	24.18	0.0
VectorRAG					
GPT-4o-mini	8 段落	0.717	0.960	17.69	0.00184
Llama-3.1-70B	8 段落	0.717	0.960	40.35	0.0

定性与专家评估

代表性查询的详细分析

为定性评估检索与推理表现，我们分析了一组具有代表性的高分子科学问题，如表 III 所示。该表并列展示了 GraphRAG 和 VectorRAG 流程针对三类不同推理问题的回答：机理理解、工艺–结构关联以及性质比较。

对于查询 1，两种流程给出的回答都符合文献记载，且在化学层面准确无误。不过，VectorRAG 对机理的解释更为详尽，具体阐述了酸性和碱性环境下的反应条件，展现出在相关信息能于段落级检索到时的优势。相比之下，GraphRAG 给出的回答简洁但不够全面；虽然内容正确，却省略了其他上下文细节，而这些细节本可丰富图元组。

对于查询 2，GraphRAG 给出了结构更清晰、层次更分明的解释，指出了生产低密度聚乳酸泡沫面临的五大类挑战，并引用了相关文献。这凸显了它在多跳推理和信息聚合方面的优势：它能将多个来源中分散的证据联系起来，形成连贯的叙述。VectorRAG 的回答虽然事实准确，但相对简短，描述深度不足。

在查询 3 中，VectorRAG 提供了 P(3HB) 和 P(4HB) 的定量热学与力学性能数据，表现优于 GraphRAG。尽管 GraphRAG 的回答因检索子图中的知识稀疏性而范围有限，但内容完全属实且没有产生幻觉，这体现了图结构支撑的推理即使在覆盖不完整时也能保持可靠性的关键优势。这再次凸显了 VectorRAG 利用段落级语义理解生成更详细、上下文更丰富的回答的能力。

总体而言，这两种流程的互补性能凸显了其不同的检索范式：GraphRAG 通过结构化推理和上下文聚合提升可解释性，同时保持事实准确性；而 VectorRAG 则擅长

在语料库中存在详细文本信息时，生成流畅的段落级解释。

表 III：GraphRAG 和 Vec- torRAG 对高分子科学代表性问题的回答定性比较，展示了两种处理流程在检索行为、推理深度和依据支撑方式上的差异。

问题	GraphRAG 回答	VectorRAG 回答
1. 什么是主要的单体产物的非生物 PHB 水解和在何种条件下它们通常是如何形成的?	非生物 PHB水解的主要单体产物是 (R)-3-羟基丁酸和巴豆酸 [4]。这一过程通常在 70–220 °C [5]的条件下进行。	非生物 PHB水解的主要单体产物是 D-(-)-3-羟基丁酸（3HBA）和巴豆酸（CA）。这些产物通常在酸性或碱性条件下形成，其中聚合物链发生水解。 具体而言，在中性 pH（7.4）下，3-羟基丁酸是聚羟基丁酸酯降解的主要中间产物降解 [1]。在酸性条件下（0.1–4.0 M H），会形成3-羟基丁酸和巴豆酸，其中巴豆酸是占主导地位的单体高浓度浓酸溶液中的产物（80–98 wt% 硫酸） [2]。相反，碱性条件（0.1–4.0 M OH-）会导致两者的产量均更高: 3-羟基丁酸和巴豆酸。
2. 什么挑战在于与.....相关生产低密度聚乳酸泡沫，以及如何这些问题是否得到了解决?	制备低密度聚乳酸泡沫面临着若干挑战，主要与材料的固有性质和加工条件有关。 1. **熔体强度和弹性：**聚乳酸表现出较低的熔体弹性和较低的熔体强度，这使得低密度泡沫的生产更加复杂。这些性能可能会妨碍在生产过程中实现所需的泡沫结构加工 [211][212]。 2. **结晶度和发泡性：**聚乳酸的高结晶度可能会限制所制泡沫的膨胀，降低其发泡性 [200]。此外，缓慢的结晶动力学会导致在获得均匀泡孔结构和理想密度方面面临的挑战 [266]。 3. **泡孔结构均匀性：**实现均匀的泡孔结构的熟妇大规模条件下尤其 难以实现低密度聚乳酸泡沫中泡孔结构的均匀性。 这种均匀性对于保持至关重要一致的力学性能贯穿整个泡沫 [213]。	制备低密度聚乳酸泡沫面临诸如聚乳酸较差的等因素的挑战熔体强度和低熔体弹性，这会导致发泡行为不佳，且未经改性。 此外，结晶行为聚乳酸的结晶行为会阻碍发泡，尤其是在结晶速率较高时，这可能会妨碍有效的泡孔形成，并生长。 CO_2 诱导的结晶在高压力下会进一步复杂化发泡过程，即通过增加降低发泡性的结晶区。 这些挑战可以通过将聚乳酸与聚羟基丁酸酯共混 （聚(3-羟基丁酸-co-3-羟基戊酸)），从而改善流变性并增强发泡性。 聚(3-羟基丁酸-co-3-羟基戊酸)可作为聚乳酸的成核剂，有望形成更好的泡孔结构并降低泡沫的密度。

	4. **表面积体积比：** 制备高 表面积体积比的泡沫是另一项 挑战，因为这会影响泡沫的整体 性能和应用 [214]。 5. **加工变量：**该 材料性质之间的相互作用 与加工条件至关重要，在 决定泡沫特性。 压力、温度以及 所用发泡剂的类型能够 显著影响最终的泡沫结构 [220][250]. 为应对这些挑战，可采用若干 策略：– **添加剂与共混：**掺入 添加剂，或将聚乳酸与其他 聚合物（如聚(3-羟基丁酸-co-3-羟基戊酸)）共混，可提高 发泡性能和机械 强度以及所得泡沫的强度 [18][130]。 **优化加工技术：** 利用先进的加工技术， 例如薄膜/泡沫共挤出，可能 提高低密度材料的生产效率 聚乳酸泡沫，从而更好地控制 加工参数 [21][264]。	调整共混物的组成 并优化加工条件，例如 发泡温度和压力，也可以 帮助获得所需的泡沫 性能，同时尽量减少不利 结晶和熔体强度 方面的局限性 [1][2][4][8]。
3. 如何 物理和 热学 的性质 P(3HB) 和 P(4HB) 有何不同 以及这些差异 意味着什么 的影响 差异 对其潜在 应用而言?	知识图谱未提供 具体的物理和 P(3HB) 和 P(4HB) 的热性能 直接提及。不过，它确实提到了 对聚羟基丁酸酯的分析，这与 P(3HB) 有关，并对热 行为和加工有影响。 1. 聚羟基丁酸酯（一种 3-羟基丁酸均聚物）通过 热分析进行分析，以 了解其特性 [4]。 2. 熔融温度与 热降解温度之间的差异会影响 聚羟基丁酸酯生产的加工性能 [20]。 3. 与其他聚合物共混可以 改善力学性能， 聚羟基脂肪酸酯的柔韧性/延展性和加工性能 [14]。	物理和热性能的 存在显著差异， 这对它们的潜在 应用具有重要影响。P(3HB)是一种高度 结晶性、水不溶性聚合物，具有 高杨氏模量（3.5 GPa）和低 拉伸强度（43 MPa）。它较脆， 断裂伸长率仅为5%， 因此不太适合需要 柔性的应用 [3]。其熔融 温度约为 180 °C，且它 的玻璃化转变温度为 约 4 °C [5]。这些 特性使其主要适用于 需要刚性和稳定性的 应用。

	尽管这些要点提供了一些背景信息 说明聚羟基丁酸酯的热性能，但它们 并未直接比较 P(3HB) 和 P(4HB)，也未提供它们具体的物理 性质及其对应用的影响。 应用。 因此，信息不足 无法从知识图谱中 全面回答关于……之间差异的问题 P(3HB) 和 P(4HB) 及其 影响。	相比之下，P(4HB) 是一种强韧且柔韧的 具有拉伸强度的热塑性塑料 可与聚乙烯媲美，并具有 100%的断裂伸长率 [4]。这种高 弹性使 P(4HB) 适用于 需要柔性和 回弹性的应用。其热性能包括 较低的熔点约为 164 °C，且 玻璃化转变温度为 3 °C [6]， 这使其更易于加工，并 能够适用于各种应用。 力学性能方面的差异 这表明，尽管 P(3HB) 可用于 刚性应用，例如包装或 结构部件，P(4HB)更适合 需要弹性的应用 和柔性的应用，例如用于生物医学 器件或柔性薄膜。此外， 能够制备共聚物，例如 P(3HB-co-4HB)，可实现定制化 特性，融合 两种聚合物的优势，适用于多种应用 [8]。

专家验证

为更好地了解检索架构如何影响推理质量，一位领域化学家使用开发的交互式RAG界面，独立评估了五种检索—生成流程的表现。分析重点考察了各系统在不同问题类型中如何平衡事实依据、上下文覆盖和引用可靠性。

比较对象包括基于 GPT-4o-mini 和 Llama-3.1-70B 构建的两个 GraphRAG 框架变体、VectorRAG、启用网页搜索的 ChatGPT-5（它集成了 OpenAI 内置的检索增强设置，将网络证据与引用追溯相结合），以及启用网页搜索的 Gemini（这是 Google 对应的 RAG 实现，可从在线来源检索支持内容）。我们使用三级问题集对每个模型进行了测试，该问题集分为通用、论文特定和多论文三个层级，旨在反映科学推理日益增加的复杂程度：从概念澄清，到对单项研究的解读，最后到对多篇出版物的综合分析。问题集的详细信息见

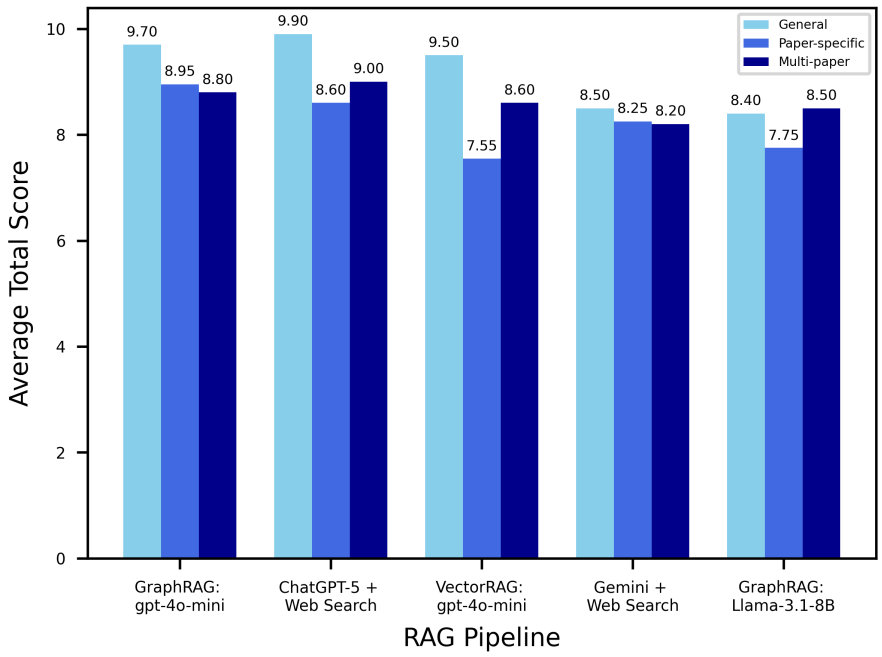

图 4. 领域专家对五种检索增强生成流程在通用、论文特定和多篇论文问题上的评估。评分反映了各系统在事实依据、上下文覆盖度和引用可靠性之间的平衡能力。GraphRAG（GPT-4o-mini）和搭载网页搜索的 ChatGPT-5 整体表现最佳，其他流程的表现中等但稳定。

补充信息。对于每个回答，专家均给出 1 至 10 分，其中 5 分用于评估内容质量和事实正确性，另外 5 分用于评估引用质量和相关性。

如图 4 所示，由 GPT-4o-mini 驱动的 GraphRAG 和搭载网页搜索的 ChatGPT-5 整体得分最高，在所有问题类别中的平均得分均为九至十分。两者均表现出扎实的事实依据、连贯的推理和可靠的引用行为。基于 Llama-3.1-70B 的 GraphRAG 变体得分略低，但事实准确性始终稳定，这说明语言模型容量会影响知识图谱元组的质量和完整性，进而影响下游答案质量。VectorRAG 和搭载网页搜索的 Gemini 在所有类别中的表现也中等且稳定。补充信息中提供了具有代表性的回答的详细比较。

在这些结果的基础上，专家评估还揭示了各系统在推理文献时存在明显且一致的差异。GraphRAG，尤其是与

GPT-4o-mini 搭配使用时，给出的回答最为准确，且有充分依据。其回答紧扣提取出的关系证据，幻觉率极低；这一特性对于科学应用至关重要。借助网络检索的 ChatGPT-5 生成了内容更丰富的叙述式回答，这反映出它能访问范围更广的信息，但其引文有时不如内部系统生成的引文具体。对于依赖段落级上下文的问题，VectorRAG 表现良好；当所需信息集中在少数几个段落中时，它往往能提供更详尽的解释。Gemini 的覆盖面更广，但由于检索到的网络来源质量参差不齐，其引用的可靠性较低。GraphRAG 搭配 Llama-3.1-70B 时，回答虽然更贴近原文，但依然准确且依据充分，这表明即使采用规模较小的开放权重模型，该框架仍然有效。

另一项重要观察是，当所提取的知识所含信息不足以回答问题时，内部流程通常会回答“我不知道”。这一行为与商业系统形成对比，后者更倾向于给出推测性或不完整的答案。宁可弃答也不产生幻觉的倾向，对科学应用尤为宝贵，因为错误论断可能误导后续解读。

综合来看，这些结果表明，建立在精心整理的知识库与领域感知加工之上的检索流程，能够匹敌并在若干方面超越网络驱动的专有系统的性能。与此同时，它们提供了更高程度的透明度、科学严谨性，以及显著更低的计算成本。与常常返回整篇文章链接、使用户必须手动定位支撑证据的商业界面不同，本文开发的流程在单个段落或知识图谱三元组层面提供精确的证据链。这降低了核验所需的工作量，增强了对生成回答的信任，并为研究人员解读和分析聚合物领域文献提供了更可靠的基础。

结论

大语言模型的最新进展展现出在科学知识发现方面的巨大潜力，但其实际效用最终取决于所依赖信息的质量与结构

。本文表明，聚合物材料领域的可靠文献研究助手并不需要庞大的专有系统或不透明的检索引擎。相反，经过精心整理的语料库结合具备领域认知的检索流程，便可构建一个既可信又贴合材料研究人员所需科学问题的基础。

通过将聚羟基脂肪酸酯作为一个具有代表性的案例进行研究，我们展示了如何调整最先进的检索方法，以体现聚合物文献的结构和报告惯例。通过 VectorRAG 实现的密集语义检索，以及通过 GraphRAG 实现的结构化关系检索，都受益于精心构建语料库、系统地转换文本以及稳健地规范化实体；这些做法共同为大语言模型在异构科学内容上进行可靠推理奠定了基础。这些设计选择使最终生成的回答可追溯、可解释，并具有科学可信度。

结果还揭示了两种检索流程各自优势的分布。VectorRAG 能够广泛进行语义召回，并提供丰富的段落级上下文，因此适用于描述性或探索性问题。GraphRAG 则擅长多跳关系推理，并能提供简洁、可解释的证据链。在语料库规模较大时，GraphRAG 能保持更高的召回率和更低的延迟；而当相关信息集中在文本某处时，VectorRAG 通常能提供更详尽的叙述式回答。两者共同彰显了在单一工作流中结合互补检索范式的价值。

在受控基准测试和专家评审中，两种流程都取得了较高的准确率。尤其是 GraphRAG，在领域问题上的表现可与商业联网检索增强生成系统相媲美，同时展现出更为谨慎、以证据为依据的行为，且引文的一致性更高。具体而言，当现有文献无法支持明确答案时，内部流程更常选择回答“我不知道”，从而降低提出推测性或缺乏依据主张的风险。与此同时，研究表明，召回率等传统检索指标无法全面反映 RAG 系统的科学实用性。召回率较低并不一定意味着依据支撑薄弱或相关性有限。判断检索到的上下文是否真正相关、推理链在科学上是否严谨，以及引文是否完整可信，仍离不开人类专业知识。

综上所述，本研究为开发经济高效且透明、并能满足材料科学界需求的人工智能学者，提供了一条实用且可复现的路径。

通过以精心整理的证据为模型推理提供依据支撑，并呈现精确到段落层面或关系层面的证据链，聚合物文献学者框架增强了人们对人工智能生成的响应的信任，并能够从日益复杂的聚合物文献中可靠地综合知识。尽管本文以 PHAs 为例进行了验证，但这种方法也广泛适用于其他材料领域，因为这些领域的科学洞见取决于对零散且异质的证据进行整合。

方法

语料库。 我们的文献语料库包含 ~ 3 百万篇文献，来源包括爱思唯尔、威利、施普林格·自然、美国化学会和英国皇家化学学会，涵盖截至2025年发表的文章。解析流程的详细信息见 [5]。

为构建PHA专用语料库，我们使用补充信息中列出的PHA相关术语，对所有文章的标题和摘要应用了基于关键词的筛选器。随后对筛选结果进行了人工核验，最终得到1,028篇PHA相关论文。之后，从这些文章的全文中提取段落，用于后续处理。

评估指标。 我们使用Recall@K、Recall PID@K和准确率评估检索性能。Recall@K是信息检索中的标准指标，用于衡量检索结果的准确率和排序质量。它量化了预期的真实上下文段落是否出现在前-K 个检索结果中。

在科学文献中，相关信息通常分布在同一篇论文的多个段落中。因此，当检索到正确论文但未检索到确切的真实段落时，Recall@K可能会低估检索性能。为解决这一局限性，我们还报告了Recall PID@K；如果前-K 个检索到的上下文中有任何一个来源于真实论文，则无论匹配的是哪个具体段落，均视为检索正确。

由于 Recall@K 和 Recall PID@K 仅关注检索行为，因此仍可能低估 RAG系统 的实际效用。因此，我们还报告准确率，该指标由领域专家评估后赋予二元（0/1）分数，评估内容为

正　　生成答案相对于问题的准确性和完整性　　ry.

Recall@K。 Recall@K 衡量预期的真实段落 P_{expected}是否包含在排名前K 的检索段落 $P_{\text{retrieved}}$中：

$$\text{Recall@K} = \begin{cases} 1, & \text{if } P_{\text{expected}} \in P_{\text{retrieved}}[:K], \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

对于 N 个查询，平均 Recall@K 为：

$$\text{Mean Recall@K} = \frac{1}{N}\sum_{i=1}^{N}\text{Recall@K}_i. \tag{2}$$

Recall PID@K。 Recall PID@K 衡量包含真实段落的论文（由其段落 ID（PID $_{\text{expected}}$）标识）是否包含在排名前K的检索论文（$\text{PID}_{\text{retrieved}}$）中：

$$\text{Recall PID@K} = \begin{cases} 1, & \text{if } \text{PID}_{\text{expected}} \in \text{PID}_{\text{retrieved}}[:K], \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

平均 Recall PID@K 的计算方式与公式（2）类似。

知识图谱构建阶段

实体抽取。通过上下文学习，利用大语言模型从语料库中抽取知识图谱三元组。每个抽取出的元组包含五个字段：*(*主语、关系、宾语、引用_关系、引用_节点*)*。如果源文本中包含显式引用和图表引用，也会一并记录。用于生成元组的优化提示词见补充信息。

实体规范化。 为统一提取出的元组中语义等价的实体，我们使用 SentenceTransformers 库中的 `all-MiniLM-L6-v2`模型生成了 384 维句子嵌入。我们采用两阶段混合聚类策略，以兼顾可扩展性和精确率。首先，使用 MiniBatchKMeans 将实体粗略分为 2,000 个聚类。随后，在每个聚类中，采用平均链接的 Agglomera-tiveClustering，并基于 0.5 的余弦距离阈值进行细粒度合并。

对于每个生成的聚类，使用基于质心的排序来选择规范实体，即选取距离聚类质心最近的实体作为代表性标签。对于数值实体占主导地位的聚类（定义为数值内容超过 60 % 的聚类），则分配形式为 `numerical_value` 的合成规范标签。此规范化步骤在保留语义关系的同时减少了词汇碎片化，从而提升了图连通性和检索鲁棒性。

GraphRAG 检索阶段

查询预处理。 在检索之前，用户查询会经过预处理，以提取领域相关术语并规范文本变体。首先将查询转换为小写，并使用 NLTK 框架实现的分词和词形还原方法进行处理。随后应用自定义正则表达式模式，识别领域特定实体，包括聚合物名称、化学式、复合材料表示法和数值。移除停用词，同时保留具有科学意义的术语。此预处理可规范查询表示，并有助于与知识图谱中的实体进行稳健匹配。

基于字符串的检索。 基于字符串的检索通过精确的关键词匹配来识别候选知识图谱三元组。对每个元组的主语、关系和宾语字段应用带有单词边界约束的正则表达式，并以不区分大小写的方式进行匹配。优先处理包含所有查询关键词的元组。如果未找到此类元组，则采用宽松匹配策略，要求至少匹配两个关键词。检索到的元组根据频次评分方案进行排序，该方案优先考虑匹配更多查询词的元组。

基于规范实体的检索。 为应对文献中的词汇变异，在实体规范化步骤中，使用生成的规范实体嵌入执行语义检索。查询词使用 `all-MiniLM-L6-v2` 模型进行嵌入，并通过余弦相似度与预先计算的规范实体嵌入进行比较。保留相似度阈值超过 0.7 的规范实体，并检索包含这些实体的所有元组。对规范实体集合进行穷尽式相似度比较

可确保涵盖语义相关的元组，这些元组可能无法仅通过基于字符串的匹配检索到。

混合检索与重排序。 混合检索结合基于字符串的匹配与基于规范形式的匹配，同时利用精确的关键词重叠和语义相似度。对于每个元组，混合分数 S_{hybrid} 通过对字符串匹配分数 S_{string} 和规范形式匹配分数 $S_{\text{canonical}}$进行加权组合计算得出：

$$S_{\text{hybrid}} = \alpha \cdot S_{\text{canonical}} + (1 - \alpha) \cdot S_{\text{string}}, \tag{4}$$

其中 $\alpha = 0.7$，体现了对语义相似度的更高重视。混合分数经过最小-最大归一化后，低于阈值 $\tau = 0.6$ 的元组（ S_{hybrid} ）将被过滤掉。

路径重排序。 经过混合匹配后保留的候选元组，将使用交叉编码器模型 `cross-encoder/ms-marco-MiniLM-L-6-v2`重新排序。该模型通过联合编码每个查询-元组对来计算重新排序分数 S_{rerank}。与实体级匹配不同，这种路径级评估会结合查询上下文，考量完整的关系路径（主语、关系、宾语），使模型能够评估整个元组的语义连贯性和相关性。此步骤通过优先排列能够更直接、全面地回答查询的元组，提高检索精确率。

最终排序分数的计算方式如下：

$$S_{\text{final}} = \lambda \cdot S_{\text{rerank}} + (1 - \lambda) \cdot S_{\text{hybrid}}, \tag{5}$$

其中， $\lambda = 0.7$ 使最终排序更倾向于上下文感知的重新排序分数。

返回元组的数量由 `max_tuples` 参数控制。为避免过早丢弃可能相关的候选项，交叉编码器最多会评估 `max_tuples`的四倍数量，将生成的分数归一化到 $[0, 1]$, 范围，并保留排名前`max_tuples` 的元组。我们评估了 `max_tuples`在 300 至 500 之间的取值，发现 300 在检索质量、计算成本和推理延迟之间取得了最佳平衡。

VectorRAG

嵌入模型。 VectorRAG 使用 SentenceTransformers 库中的 `Qwen/Qwen3-Embedding-4B` 模型，为语料库和用户查询生成 2,560 维嵌入。语料库嵌入会预先计算，并存储在 PostgreSQL 数据库中，以便高效检索。

文本分块策略。 科学文章通常分为引言、方法、结果与讨论、结论等主要章节，每个章节可能包含多个层级小节。为在检索过程中保留科学语境，我们根据这种结构层级合并段落。具体而言，将属于同一个一级小节（即主要章节下级别最低的共同子标题）的所有段落合并为一个文本块。这种方法在优化语义理解和检索性能的同时，保留了章节内科学论证的逻辑脉络与连贯性。

数据可用性

由于本研究未创建或分析任何新数据，因此本文不适用数据共享。

代码可用性

所用代码　　本研究内容可见于 https://github.com/Ramprasad-Group/RAG

致谢

本研究由美国海军研究办公室通过拨款 N00014-19-1- 2103and N00014-20-1-2175 资助。

利益冲突

作者声明不存在利益冲突。

补充信息

本文的在线版本包含可在 https:获取的补充信息。

//doi.org/

- VectorRAG和GraphRAG组件的优化提示词、大语言模型基准测试、领域专家评估问题等。

参考文献

[1] A. Agrawal 和 A. Choudhary，观点：材料信息学与大数据：在材料科学中实现科学的“第四范式”，Apl Materials 4 (2016)。[2] R. Ramprasad、R. Batra、G. Pilania、A. Mannodi-Kanakkithodi 和 C. Kim，材料信息学中的机器学习：近期应用与前景，npj Computational Materials3，54 (2017)。[3] L. Himanen、A. Geurts、A. S. Foster 和 P. Rinke，数据驱动的材料科学：现状、挑战与展望，Advanced Science6，1900808 (2019)。[4] T. B. Martin 和 D. J. Audus，机器学习的新兴趋势：聚合物视角，ACS Polymers Au 3，239 (2023)。[5] S. Gupta、A. Mahmood、P. Shetty、A. Adeboye 和 R. Ramprasad，使用大语言模型从聚合物文献中抽取数据，Communications materials 5，269 (2024)。[6] P. Shetty、A. C. Rajan、C. Kuenneth、S. Gupta、L. P. Panchumarti、L. Holm、C. Zhang 和 R. Ramprasad，使用自然语言处理从大型聚合物语料库中抽取材料属性数据的通用处理流程，npj Comput Mater 9，1 (2023)。[7] J. Cheung、Y. Zhuang、Y. Li、P. Shetty、W. Zhao、S. Grampurohit、R. Ramprasad 和 C. Zhang，Polyie：一个从聚合物材料科学文献中进行信息抽取的数据集，载于*2024*年北美计算语言学协会分会会议论文集：人类语言技术（第*1*卷：长篇论文）(2024) 第2370–2385页。

[8] P. Kalhor, N. Jung, S. Bräse, C. Wöll, M. Tsotsalas 和 P. Friederich，《自动化数据抽取与机器学习赋能的功能材料系统》，Advanced Functional Materials34, 2302630(2024)。

[9] F. Bai, J. Kang, G. Stanovsky, D. Freitag, M. Dredze 和 A. Ritter, 《从异构表格中进行模式驱动的信息抽取》, 载于计算语言学协会年会论文集：EMNLP 2024 (2024), 第 10252–10273 页。[10] S. Gupta, A. Mahmood, S. Shukla 和 R. Ramprasad, 《大语言模型聚合物性能预测基准测试》, 《大分子快速通讯》, e00388 (2025)。[11] K. Stergiou, C. Ntakolia, P. Varytis, E. Koumoulos, P. Karlsson 和 S. Moustakidis, 《通过机器学习提升建筑材料的性能预测与工艺优化：综述》, 《计算材料科学》220, 112031 (2023)。[12] L. Chen, J. Kern, J. P. Lightstone 和 R. Ramprasad, 《数据辅助的聚合物逆合成规划》, 《应用物理评论》8, 031405 (2021)。[13] E. Kim, K. Huang, A. Saunders, A. McCallum, G. Ceder 和 E. Olivetti, 《通过文本抽取与机器学习从科学文献中获取材料合成见解》, 《材料化学》29, 9436 (2017)。[14] A. Savit, H. Sahu, S. Shukla, W. Xiong 和 R. Ramprasad, 《polybart：用于聚合物性能预测与生成式设计的化学语言模型》, arXiv 预印本 arXiv:2506.04233 (2025)。[15] H. Sahu, W. Xiong, A. Savit, S. S. Shukla 和 R. Ramprasad, 《用于生成式聚合物设计的编码器—解码器基础化学语言模型》, arXiv 预印本 arXiv:2510.18860 (2025)。[16] I. Augenstein, T. Baldwin, M. Cha, T. Chakraborty, G. L. Ciampaglia, D. Corney, R. DiResta, E. Ferrara, S. Hale, A. Halevy 等, 《大语言模型时代的事实性挑战与事实核查机遇》, 《自然·机器智能》6, 852 (2024)。[17] A. Alansari 和 H. Luqman, 《大语言模型幻觉：全面综述》(2025), arXiv:2510.06265 [cs.CL].[18] A. T. Kalai, O. Nachum, S. S. Vempala 和 E. Zhang, 《语言模型为何会产生幻觉》(2025), arXiv:2509.04664 [cs.CL].[19] S. M. T. I. Tonmoy, S. M. M. Zaman, V. Jain, A. Rani, V. Rawte, A. Chadha 和 A. Das, 《大语言模型幻觉缓解技术全面综述》(2024), arXiv:2401.01313 [cs.CL].[20] J.-W. Shim, Y.-J. Ju, J.-H. Park 和 S.-W. Lee, 《用于缓解大语言模型幻觉的多阶段提示优化》(2025), arXiv:2510.12032 [cs.CL].[21] A. Madaan, N. Tandon, P. Gupta, S. Hallinan, L. Gao, S. Wiegreffe, U. Alon, N. Dziri, S. Prab-

humoye、Y. Yang、S. Gupta、B. P. Majumder、K. Hermann、S. Welleck、A. Yazdanbakhsh和P. Clark，自我精炼：基于自我反馈的迭代精炼（2023年），arXiv:2303.17651 [cs.CL].[22]O. Ayala和P. Bechard，通过检索增强生成降低结构化输出中的幻觉，载于*2024年北美计算语言学协会分会会议论文集：人类语言技术（第6卷：工业赛道）*（计算语言学协会，2024年）第228–238页。[23] W. Fan、Y. Ding、L. Ning、S. Wang、H. Li、D. Yin、T.-S. Chua和Q. Li，检索增强生成遇见大语言模型综述：迈向检索增强大语言模型（2024年），arXiv:2405.06211[cs.CL].[24] S. Ahmad、Z. Nezami、M. Hafeez和S. A. R. Zaidi，面向开放式无线接入网（ORAN）的向量、图与混合检索增强生成（检索增强生成）流程基准测试，arXiv预印本 arXiv:2507.03608（2025年）。[25] S. Kukreja、T. Kumar、V. Bharate、A. Purohit、A. Dasgupta和D. Guha，结合检索增强生成的向量嵌入性能评估，载于*2024年第9届国际计算机与通信系统会议（ICCCS）*（2024年）第333–340页。[26] B. Peng、Y. Zhu、Y. Liu、X. Bo、H. Shi、C. Hong、Y. Zhang和S. Tang，图检索增强生成：综述（2024年），arXiv:2408.08921 [cs.AI].[27]K. Olonisakin、A. K. Mohanty、M. Thimmanagari和M. Misra，可生物降解聚合物共混物及其生物复合材料的最新进展：综合综述，《绿色化学》**27**，11656（2025年）。