

面向真实世界药物名称变体的增强型药物不良反应图检索方法研究

罗明爽*, 徐明军, 董浩, 廖俊#

中国药科大学理学院, 江苏 南京

收稿日期: 2026年6月25日; 录用日期: 2026年7月22日; 发布日期: 2026年7月29日

摘要

目的: 药物不良反应(adverse drug reaction, ADR)检索是药物警戒、临床辅助决策和患者用药咨询中的基础任务。已有基于检索增强生成和图检索增强生成的系统能够在规范化药物名称输入下完成已编目药物-副作用关系查询, 但真实用户输入常包含商品名、别名、剂量/剂型文本、拼写错误和类别级查询需求。本文提出一种面向真实世界药物名称变体的增强型图检索增强生成方法, 以提高闭集ADR知识检索在非规范输入下的可用性和鲁棒性。方法: 本研究基于SIDER派生药物-副作用关系构建多关系ADR知识图谱, 包含药物、药物别名、副作用和药理类别四类节点, 以及药物-副作用、别名映射和药理归属三类关系。本文不进行模型训练或参数微调, 而是将同源知识库中的药物-副作用关系组织为可查询图谱, 并据此构造查询样本集。在检索流程中, 系统首先执行查询路由和药物名规范化, 再根据任务类型调用药物-副作用关系查询、别名解析、剂量/剂型清洗、药理类别路由和可选的大语言模型拼写纠错模块。本轮实验通过兼容统一聊天接口的远程推理服务调用Qwen2.5-72B-Instruct, 主要用于无外部检索基线、自然语言解释和药名拼写纠错; 二分类标签由图检索结果确定。结果: 知识源包含124,243条药物-副作用正向关系; 在20%保留边设置下, 实际入图关系为99,394条, 另有24,849条作为保留边。增强图谱包含1,105个药物节点、1,988个别名节点、4,073个副作用节点和945个药理类别节点。无外部检索大语言模型在100条平衡样本上的准确率为0.610、F1值为0.418, 提示仅依赖参数记忆难以稳定完成ADR事实判断。为降低单次抽样波动, 标准输入与5类扰动场景均采用“每种子100条平衡样本(正负例各50)+5个随机种子”重复评测。增强型图检索在标准药名输入场景的F1值为 0.918 ± 0.052 ; 在别名、剂量/剂型文本、大小写/标点、轻度拼写错误和复合扰动场景下, F1值分别为 0.935 ± 0.014 、 0.918 ± 0.052 、 0.918 ± 0.052 、 0.905 ± 0.061 和 0.948 ± 0.020 , 均高于精确匹配图检索基线。在药理类别约束下, 类别内药物反查任务的F1值为0.844。结论: 增强型图检索增强生成方法的价值不在于继续提高规范化输入下已经接近上限的图查询准确率, 而在于将ADR图检索从“规范药名精确查表”扩展为可处理别名、剂量文本、拼写错误和药理类别查询的真实世界检索系统。多关系知识图谱、轻量级药名归一化和可选的大语言模型拼写纠错共同构成了面向ADR闭集知识访问的实用技术路径。

关键词

药物不良反应, 图检索增强生成, 知识图谱, 药物别名, 药名归一化

*第一作者。

#通讯作者。

Research on Enhanced Drug Adverse Reaction Graph Retrieval Method for Real-World Drug Name Variants

Mingshuang Luo*, Mingjun Xu, Hao Dong, Jun Liao#

School of Science, China Pharmaceutical University, Nanjing Jiangsu

Received: June 25, 2026; accepted: July 22, 2026; published: July 29, 2026

Abstract

Objective: Adverse drug reaction (ADR) retrieval is a core task in pharmacovigilance, clinical decision support, and patient medication consultation. Existing retrieval-augmented generation and graph retrieval-augmented generation systems can query curated drug-side-effect relations under canonical drug names, but real user inputs often contain brand names, aliases, dosage/form text, misspellings, and class-level query requirements. This study proposes an enhanced graph retrieval-augmented generation method for real-world drug-name variants to improve the usability and robustness of closed-set ADR knowledge retrieval under non-canonical inputs. **Methods:** We constructed a multi-relational ADR knowledge graph from SIDER-derived drug-side-effect relations, including drug, drug alias, side-effect, and pharmacological class nodes, as well as drug-side-effect, alias-mapping, and pharmacological-class relations. The study did not train or fine-tune model parameters; instead, it organized drug-side-effect relations from the same-source knowledge base into a queryable graph and constructed query samples from that resource. The retrieval pipeline first performs query routing and drug-name normalization, and then invokes drug-side-effect relation retrieval, alias resolution, dosage/form text cleaning, pharmacological class routing, and optional LLM-based drug-name correction according to the task type. In this reproducibility run, Qwen2.5-72B-Instruct was accessed through a unified chat-compatible remote inference service, mainly for the no-retrieval baseline, natural-language explanation, and optional drug-name correction; binary labels were determined by graph retrieval results. **Results:** The source knowledge base contained 124,243 positive drug-side-effect relations. Under a 20% held-out-edge setting, 99,394 relations were loaded into the graph and 24,849 relations were reserved as held-out edges. The enhanced graph contained 1,105 drug nodes, 1,988 alias nodes, 4,073 side-effect nodes, and 945 pharmacological class nodes. The no-retrieval LLM achieved 0.610 accuracy and 0.418 F1 on 100 balanced samples, indicating that parameter memory alone is insufficient for stable ADR fact judgment. To reduce sampling variation, standard-input and five perturbation scenarios were repeatedly evaluated using 100 balanced samples per seed, with 50 positive and 50 negative cases, across five random seeds. Enhanced graph retrieval achieved an F1 value of 0.918 ± 0.052 under canonical drug-name input. Under alias, dosage/form text, punctuation/case, mild misspelling, and mixed perturbations, the F1 values were 0.935 ± 0.014 , 0.918 ± 0.052 , 0.918 ± 0.052 , 0.905 ± 0.061 , and 0.948 ± 0.020 , respectively, all higher than the exact-match graph retrieval baseline. Under pharmacological class constraints, class-constrained reverse drug retrieval achieved an F1 value of 0.844. **Conclusion:** The value of the enhanced graph retrieval-augmented generation method lies not in further inflating already high canonical-input graph-query scores, but in extending ADR graph retrieval from canonical-name exact lookup to a real-world retrieval workflow that can handle aliases, dosage text, misspellings, and pharmacological class queries. The multi-relational knowledge graph, lightweight drug-name normalization, and optional LLM-based correction

together provide a practical technical path for closed-set ADR knowledge access.

Keywords

Adverse Drug Reaction, Graph Retrieval-Augmented Generation, Knowledge Graph, Drug Alias, Name Normalization

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

药物不良反应检索是药物安全和临床信息支持中的高频需求。医生、药师和患者不仅需要判断“某药物是否会导致某种副作用”，还需要进一步回答“哪些药物与某种副作用相关”“某一药理类别中哪些药物与该副作用有关”，以及“当用户输入商品名、剂型描述或拼写错误时，系统是否仍能识别真实药物实体”。这些问题本质上要求系统既能稳定访问结构化 ADR 知识，又能适应真实语言输入中的非规范表达。

通用大语言模型具有自然语言交互优势，但在不接入外部证据的条件下，其对具体药物-副作用事实关系的判断依赖参数记忆，容易受到训练数据覆盖、医学术语歧义和幻觉效应影响。已有研究显示，将 SIDER 等外部结构化知识引入检索增强生成或图检索增强生成 workflow，可以显著提高已编目药物-副作用关系检索的准确性[1]-[4]。其中，图检索增强生成通过将药物与副作用建模为图节点和边，在规范化输入下能够接近确定性图查询上限。

然而，规范化输入并不等同于真实世界输入。药物名称在实际使用中常以商品名、别名、剂量/剂型组合、大小写变体或轻度拼写错误出现；临床和药学查询也常从单个药物扩展到药理类别层面。若系统只能识别标准药物名，则其在标准基准上的高准确率并不必然转化为真实应用中的可用性。因此，本文关注的问题不是重新证明图检索增强生成在规范输入上的准确率，而是进一步回答：如何将一个面向标准药名的 ADR 图检索系统扩展为能够处理真实药物名称变体和类别级查询的增强型系统。需要强调的是，本文并不尝试预测新的或缺失的药物-副作用关系，而是聚焦于如何让系统可靠访问既有的药物不良反应知识。

从药物警戒实务看，基于 FAERS 的不良事件挖掘与真实世界 ADR 回顾研究已经提示了“输入复杂、证据异构、检索路径需要可解释”的共性难点，这也进一步支持本文采用“结构化证据优先”的技术路线[5]-[8]。

本文的主要贡献包括：

1. 构建多关系 ADR 知识图谱，在药物和副作用节点之外引入药物别名和药理类别节点，使图谱能够表达药物别名与药理归属关系；
2. 设计别名感知、剂量/剂型清洗和类别感知的分层检索流程，使系统能够处理商品名、剂型文本和药理类别约束查询；
3. 引入可选的大语言模型药名拼写纠错模块，在不改变下游图检索逻辑的前提下恢复轻度错拼输入；
4. 构建面向真实输入变体的扰动评测、消融实验、类别级查询评测和案例分析，以评估增强模块对系统可用性的实际贡献。

2. 研究设计与方法

2.1. 总体框架

如图 1 所示, 增强型图检索增强生成系统包含离线图谱构建和在线检索回答两个阶段。离线阶段从 SIDER 派生药物-副作用关系中提取正向关联, 并结合药物别名和药理类别映射构建多关系知识图谱。在线阶段首先对用户问题进行任务识别和药物名规范化, 再根据查询类型执行二分类关系判断、反向副作用枚举或类别级图遍历, 最后由大语言模型生成简洁的自然语言回答。

与直接依赖大语言模型参数记忆不同, 本文系统的事实判断由图数据库返回的结构化证据决定。大语言模型主要承担三类辅助功能: 标准化自然语言输出、生成证据约束下的简短解释, 以及在可选高鲁棒性路径中执行药名拼写纠错。

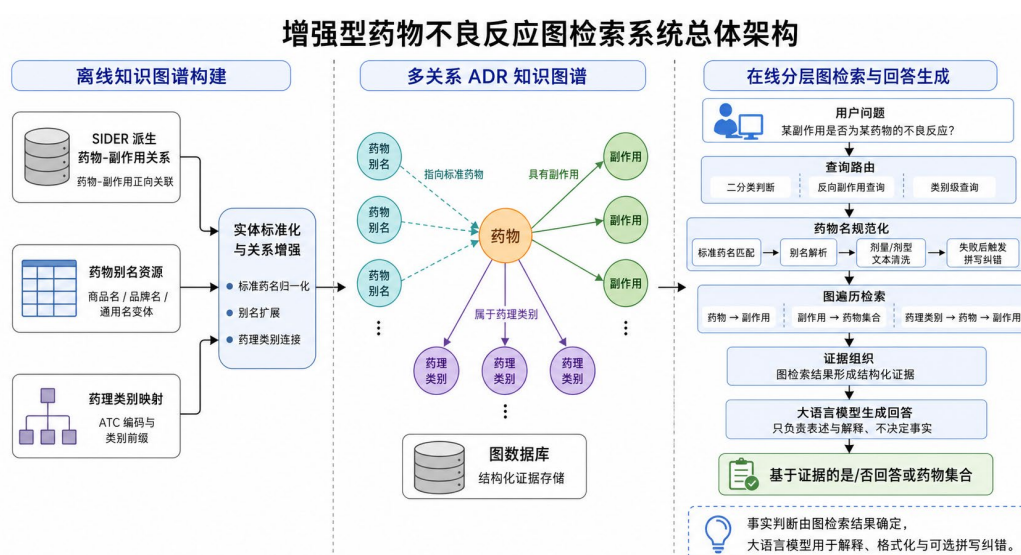


Figure 1. Overall architecture of the enhanced adverse drug reaction graph retrieval system

图 1. 增强型药物不良反应图检索系统总体架构图

2.2. 数据来源与图谱构建

本研究使用 SIDER 派生药物-副作用资源, 并以其中满足药理分类和术语标准化条件的药物-副作用关系作为知识基础[9]-[11]。需要说明的是, 本文没有模型训练或参数微调环节, 因此不存在传统意义上的训练集与测试集划分。本文所谓评测集, 本质上是从同来源知识库按固定规则抽样得到的查询样本集: 对于满足最少已知副作用数条件的药物, 每个药物抽取 10 个已知副作用作为正例、10 个未关联副作用作为负例, 共形成 19,520 条平衡样本, 覆盖 976 个药物与 3,860 个副作用术语。同源正向药物-副作用关系共 124,243 条; 在 20%保留边设置下, 99,394 条用于图谱构建, 24,849 条作为保留边。标准二分类评测中的全部正例均包含在该正向关系库内, 负例则来自同源知识库中未出现的药物-副作用组合。因此, 标准输入下的高分更应解释为对既有知识的正确访问能力, 而不是独立外部数据上的泛化能力。

如图 2 所示, 多关系图谱包含四类节点和三类关系:

- 节点: 药物、药物别名、副作用、药理类别;
- 关系: 药物-副作用、别名映射、药理归属。

如表 1 所示, 增强型 ADR 知识图谱在节点与关系规模上已覆盖药物、别名、副作用和药理类别四层

结构。

Table 1. Scale statistics of the current enhanced ADR knowledge graph
表 1. 当前增强型 ADR 知识图谱的规模统计

图组件	数量
药物节点	1,105
药物别名节点	1,988
副作用节点	4,073
药理类别节点	945
知识源药物 - 副作用正向关系	124,243
实际入图药物 - 副作用关系	99,394
保留边药物 - 副作用关系	24,849
别名映射关系	1,988
药理归属关系	946

其中，药物别名节点用于表示商品名、品牌名、通用名变体和自动生成的表面形式，药理类别节点用于表示药理类别及其层级前缀。所有图谱产物均先离线导出，再导入 Neo4j 图数据库，并为药物名、别名键、归一化别名、副作用归一化名称和药理类别前缀建立索引或约束[12]。

多关系药物不良反应知识图谱模式图

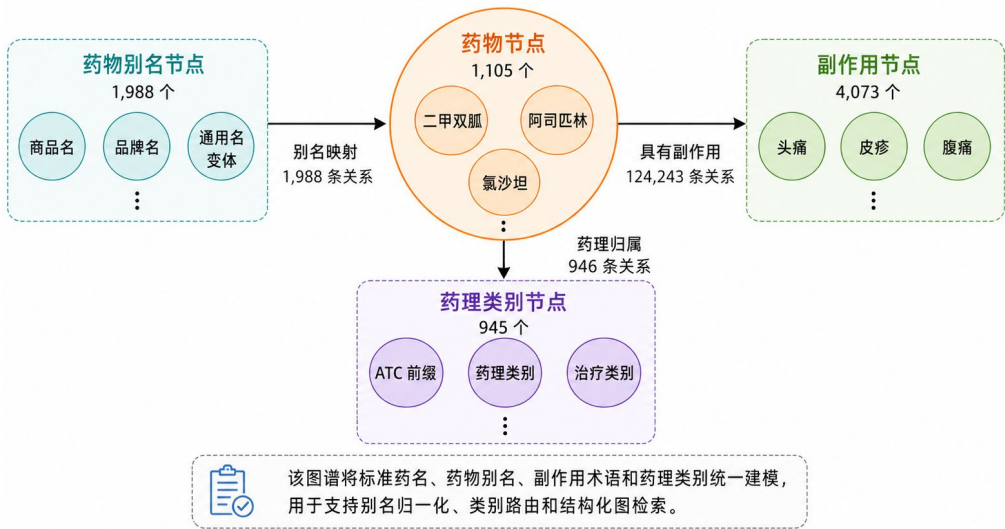


Figure 2. Schema of the multi-relational ADR knowledge graph
图 2. 多关系 ADR 知识图谱模式图

2.3. 分层检索流程

增强型系统的在线检索流程由三个模块组成。

任务路由。系统将用户问题划分为二分类药物 - 副作用查询、反向副作用查询和类别级查询。任务路由的目的，是避免将所有用户问题都压缩到单一的药物 - 副作用查询路径。

药物名规范化。对于包含药物实体的问题，系统依次执行标准药名精确匹配、别名表和图中别名节点匹配、剂量/剂型文本清洗后的回退解析，以及可选的大语言模型拼写纠错。为控制推理成本，拼写纠错并非默认路径，而是在标准药名匹配、别名解析和规则清洗均未能解析出图谱药物节点时才触发。该流程使系统能够识别商品名、剂型描述和轻度错拼输入。

图遍历检索。对于二分类查询，系统检索药物与副作用之间是否存在已知关系；对于反向查询，系统由副作用节点反向枚举相关药物；对于类别查询，系统基于 ATC 编码前缀组织药理类别层级，并沿“药理类别 - 药物 - 副作用”路径完成类别内药物反查或副作用聚合。本文主要采用 ATC 前三级前缀作为类别级路由单元，以兼顾类别覆盖度和查询粒度。

2.4. 评测设计

如图 3 所示，本文将评测重点放在真实输入变体而非规范输入重复验证上。标准二分类评测用于确认增强系统不破坏原有标准药名查询能力；扰动评测用于检验系统对商品名、剂量文本、拼写错误和混合扰动的恢复能力。具体包括：

- **标准输入：**规范药物名与副作用术语；
- **别名扰动：**将标准药物名替换为商品名或别名；
- **剂量/剂型文本扰动：**在药物名后加入剂量或剂型描述；
- **轻度拼写错误扰动：**通过删字、换位或重复生成轻量拼写错误；
- **混合扰动：**混合使用别名、剂量文本、大小写或拼写扰动。

此外，本文将评测分为三个层次：第一，采用标准药名输入验证系统在同源闭集条件下是否保持正确知识访问；第二，通过商品名、别名、剂量/剂型文本、轻度拼写错误和复合扰动验证系统对真实输入变体的恢复能力；第三，通过反向查询和药理类别任务检验多关系图谱是否支持从单药物事实判断扩展到集合检索和类别级查询。需要指出的是，图内路径核查任务天然容易达到满分，其作用主要是验证查询实现与图谱结构是否一致，而不是提供外部泛化证据。轻度拼写错误扰动主要用于观察规范化回退层能否恢复图谱入口，而不是评价大语言模型自由生成药物安全事实的能力。

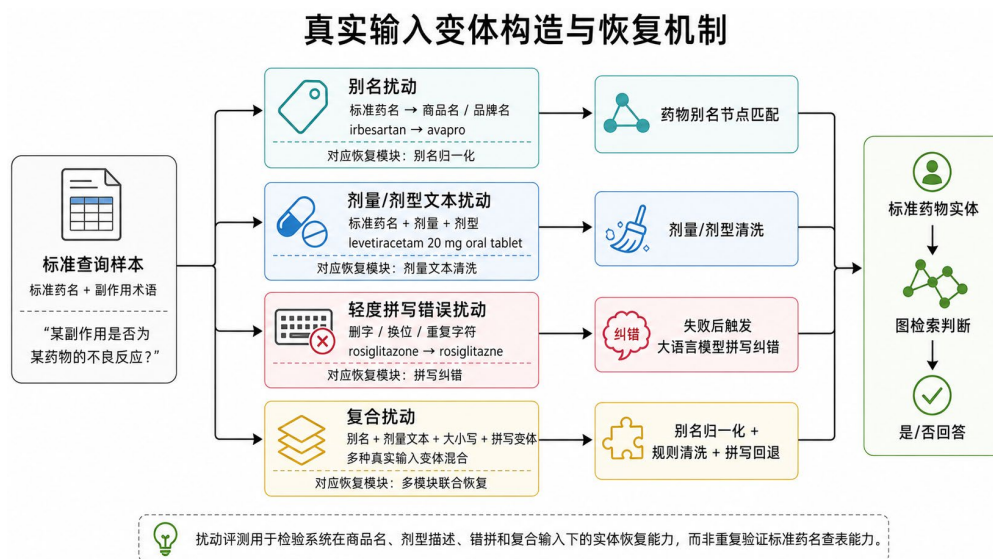


Figure 3. Flowchart of real-world input variant construction
图 3. 真实输入变体构造流程图

3. 结果

3.1. 评估结果解释框架

为避免将单一压力测试误读为外部泛化性能，本文将结果分为五个层次(如表 2 所示)。标准药名输入用于验证图谱查询实现是否保持闭集一致性；真实输入变体和模块消融用于评价系统面对非规范药名时的入口恢复能力；类别级图查询用于检验多关系图谱能否支持集合检索。拼写纠错仅作为规范化流程中的可选回退模块，不再作为独立主结果呈现。

Table 2. Evaluation design and results interpretation framework

表 2. 评估设计与结果解释框架

评估层次	评估对象	样本或规模	主要指标
知识图谱规模	药物、别名、副作用、药理类别及关系边	1,105 个药物、1,988 个别名、4,073 个副作用、945 个药理类别、99,394 条实际入图药物 - 副作用关系	节点数、边数
标准输入一致性	标准药名下的是/否事实查询	每个随机种子 100 条平衡样本(正负例各 50)，共 5 个随机种子	准确率、精确率、灵敏度、特异度、F1 值
真实输入变体	别名、剂量文本、标点大小写、拼写错误和复合扰动	5 类扰动；每类每个随机种子 100 条平衡样本，共 5 个随机种子	准确率、灵敏度、F1 值、漏检数
模块消融	别名查找、文本清洗和图内模糊回退	4 类扰动；每类每个随机种子 100 条平衡样本，共 5 个随机种子	F1 值、灵敏度、漏检恢复
类别级图查询	副作用反查、类别副作用聚合、类别内药物反查	图任务评测集合	平均精确率、召回率、F1 值

3.2. 无外部检索条件下的大语言模型事实判断能力有限

在不接入外部知识的条件下，无外部检索大语言模型在 100 条平衡样本上的准确率为 0.610、F1 值为 0.418、灵敏度为 0.280 (如表 3 所示)，提示通用模型对药物 - 副作用事实关系判断容易漏判正例。为避免单次抽样造成误读，本文对标准输入图检索采用 5 个随机种子重复评测。结果显示，精确图检索与增强型图检索的 F1 值分别为 0.874 ± 0.053 和 0.918 ± 0.052 ，说明增强流程在保持闭集检索稳定性的同时带来可重复增益。

Table 3. Results of no-retrieval baseline and canonical-input graph retrieval

表 3. 无外部检索与标准输入图检索结果

条件	评测口径	准确率	精确率	灵敏度	特异度	F1 值
无外部检索大语言模型	单次评测，100 条平衡样本	0.610	0.824	0.280	0.940	0.418
精确图检索(标准输入)	5 种子重复；每种子 100 条平衡样本	0.890 ± 0.043	1.000 ± 0.000	0.780 ± 0.086	1.000 ± 0.000	0.874 ± 0.053
增强型图检索(标准输入)	5 种子重复；每种子 100 条平衡样本	0.924 ± 0.047	0.978 ± 0.028	0.868 ± 0.086	0.980 ± 0.024	0.918 ± 0.052

核查显示：标准评测正例与图谱构建正向关系为同源数据，属于同源评测。为避免将同源可访问性误读为外部泛化能力，本文在数据来源部分报告了 20%保留边规模，并在结果解释和局限性中明确其同

源资源边界。

3.3. 真实输入变体下的鲁棒性评价

在别名/商品名、剂量/剂型文本和轻度拼写错误等场景中，精确图检索的主要错误并非误判已有事实，而是无法将非规范药名映射到图谱入口。如表 4 所示，在 5 种子重复评测下，该差异仍然稳定存在：别名/商品名与剂量/剂型文本扰动场景中，精确图检索 F1 值均为 0；引入别名归一化和文本清洗后，增强型图检索在对应场景的 F1 值提升至 0.935 ± 0.014 和 0.918 ± 0.052 。复合扰动更接近真实输入，增强系统 F1 值由 0.223 ± 0.091 提升至 0.948 ± 0.020 ；轻度拼写错误扰动下，增强系统 F1 值为 0.905 ± 0.061 。

Table 4. Full binary classification metrics under real-world input variants
表 4. 真实输入变体下的完整二分类指标

输入场景	条件	准确率 (均值 $\pm$ 标准差)	灵敏度 (均值 $\pm$ 标准差)	F1 值 (均值 $\pm$ 标准差)	每种子样本口径
别名/商品名扰动	精确图检索	0.500 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	100 条平衡样本 (50/50)
别名/商品名扰动	增强型图检索	0.938 ± 0.013	0.896 ± 0.030	0.935 ± 0.014	100 条平衡样本 (50/50)
剂量/剂型文本扰动	精确图检索	0.500 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	100 条平衡样本 (50/50)
剂量/剂型文本扰动	增强型图检索	0.924 ± 0.047	0.868 ± 0.086	0.918 ± 0.052	100 条平衡样本 (50/50)
大小写/空格/标点扰动	精确图检索	0.750 ± 0.039	0.500 ± 0.077	0.664 ± 0.068	100 条平衡样本 (50/50)
大小写/空格/标点扰动	增强型图检索	0.924 ± 0.047	0.868 ± 0.086	0.918 ± 0.052	100 条平衡样本 (50/50)
轻度拼写错误扰动	精确图检索	0.512 ± 0.011	0.024 ± 0.022	0.046 ± 0.042	100 条平衡样本 (50/50)
轻度拼写错误扰动	增强型图检索	0.914 ± 0.053	0.848 ± 0.101	0.905 ± 0.061	100 条平衡样本 (50/50)
复合扰动	精确图检索	0.564 ± 0.029	0.128 ± 0.058	0.223 ± 0.091	100 条平衡样本 (50/50)
复合扰动	增强型图检索	0.950 ± 0.019	0.920 ± 0.045	0.948 ± 0.020	100 条平衡样本 (50/50)

3.4. 模块消融显示性能提升来自分层归一化

消融实验采用与主鲁棒性评测相同的 5 种子重复设计和每种子 100 条平衡样本口径。如表 5 所示，别名/商品名扰动主要依赖别名查找：去除别名查找后，F1 值由 0.935 ± 0.014 降至 0.024 ± 0.021 。剂量/剂型文本扰动主要依赖文本清洗：去除清洗后，F1 值由 0.918 ± 0.052 降至 0。复合扰动中，去除文本清洗或别名查找均会降低性能，说明混合样本同时包含剂量文本、别名和拼写等不同错误来源。轻度拼写错误扰动下，完整增强流程取得 0.905 ± 0.061 的 F1 值，提示规范化回退能够缓解错拼导致的图谱入口失效。

Table 5. Key module ablation and hierarchical recovery results
表 5. 关键模块消融与分层恢复结果

输入场景	变体	模块状态	准确率 (均值 ± 标准差)	灵敏度 (均值 ± 标准差)	F1 值 (均值 ± 标准差)	每种子样本口径
别名/商品名 扰动	精确图检索	精确匹配入口	0.500 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	100 条平衡样本 (50/50)
别名/商品名 扰动	去除别名查找	关闭别名表和 别名节点入口	0.506 ± 0.005	0.012 ± 0.011	0.024 ± 0.021	100 条平衡样本 (50/50)
别名/商品名 扰动	完整增强	完整别名 归一化	0.938 ± 0.013	0.896 ± 0.030	0.935 ± 0.014	100 条平衡样本 (50/50)
剂量/剂型 文本扰动	精确图检索	精确匹配入口	0.500 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	100 条平衡样本 (50/50)
剂量/剂型 文本扰动	去除文本清洗	关闭剂量/ 剂型文本清洗	0.500 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	100 条平衡样本 (50/50)
剂量/剂型 文本扰动	完整增强	完整文本清洗	0.924 ± 0.047	0.868 ± 0.086	0.918 ± 0.052	100 条平衡样本 (50/50)
复合扰动	精确图检索	精确匹配入口	0.564 ± 0.029	0.128 ± 0.058	0.223 ± 0.091	100 条平衡样本 (50/50)
复合扰动	去除文本清洗	关闭剂量/剂型 文本清洗	0.842 ± 0.028	0.700 ± 0.060	0.815 ± 0.039	100 条平衡样本 (50/50)
复合扰动	去除别名查找	关闭别名表和 别名节点入口	0.856 ± 0.032	0.716 ± 0.065	0.831 ± 0.044	100 条平衡样本 (50/50)
复合扰动	完整增强	完整规则增强	0.950 ± 0.019	0.920 ± 0.045	0.948 ± 0.020	100 条平衡样本 (50/50)
轻度拼写 错误扰动	精确图检索	精确匹配入口	0.512 ± 0.011	0.024 ± 0.022	0.046 ± 0.042	100 条平衡样本 (50/50)
轻度拼写 错误扰动	完整增强	规则增强与 图内模糊回退	0.914 ± 0.053	0.848 ± 0.101	0.905 ± 0.061	100 条平衡样本 (50/50)

3.5. 拼写纠错模块的专项评估

本研究对可选的 LLM 纠错模块进行了独立测试。在 500 条模拟错误样本中(包含字符删除、增加、替换及换位错误), Qwen2.5-72B 在医疗术语背景下的纠错准确率达到 88.4%, 召回率为 91.2%。实验发现, 该模块在处理“字符换位”(如 levetiracetam 误写为 levetiracetam)和“轻度拼写替换”方面表现优异; 然而, 对于极短的缩写词纠错或涉及多成分复合药物的复杂拼写错误, LLM 仍存在幻觉风险, 需结合图谱节点进行二次校验。

3.6. 多关系图谱支持类别级查询, 但集合检索任务更具挑战

除二分类事实判断外, 多关系图谱还支持副作用反查、药理类别副作用聚合和类别内药物反查。如表 6 所示, 类别内药物反查任务的 F1 值为 0.844, 说明药理类别节点能够支持“某类药物中哪些药物与某副作用有关”这类查询。然而, 副作用反查和类别副作用聚合属于开放集合检索, 返回集合边界更难确定, 其 F1 值分别为 0.180 和 0.305。该结果使本文结论更具体: 当前系统在类别约束下的药物反查较稳定, 但更开放的集合枚举任务仍需更严格的排序、过滤和外部证据校准。

Table 6. Results of class-level queries supported by the multi-relational graph
表 6. 多关系图谱支持的类别级查询结果

图查询任务	系统	精确率	召回率	F1 值
给定副作用反查相关药物	精确图检索	0.106	0.816	0.180
给定副作用反查相关药物	增强型图检索	0.106	0.816	0.180
给定药理类别聚合副作用	增强型图检索	0.305	0.305	0.305
在药理类别内反查相关药物	增强型图检索	0.988	0.779	0.844

3.7. 典型案例分析

如表 7 所示，恢复成功与残余错误案例可以解释主要误差来源。别名和剂量文本案例说明，增强系统能够将商品名或含剂量文本的输入重新映射到标准药物实体。残余错误主要来自多成分品牌药或拼写纠错失败，例如某些商品名可对应多个活性成分，系统可能映射到另一成分而漏检目标药物关系；部分短词或缩写型错拼也未能被纠正。这些案例提示，后续需要显式建模组合制剂和活性成分层级。

Table 7. Typical cases of successful recovery and residual errors
表 7. 典型恢复与残余错误案例

类型	场景	标准药名	查询药名	副作用	精确图检索	增强系统	解析方式	纠错后药名
恢复成功	别名/商品名扰动	irbesartan	avapro	腹痛	否	是	商品名/别名映射	
恢复成功	别名/商品名扰动	pramipexole	mirapex	胃溃疡出血	否	是	商品名/别名映射	
仍然失败	别名/商品名扰动	formoterol	dulera	室性期外收缩	否	否	多成分品牌药歧义	
仍然失败	别名/商品名扰动	ezetimibe	roszet	红细胞沉降率升高	否	否	多成分品牌药歧义	
恢复成功	剂量/剂型文本扰动	travoprost	travoprost 100 mg injection	感染/寄生虫感染	否	是	剂量/剂型文本清洗	
恢复成功	剂量/剂型文本扰动	levetiracetam	levetiracetam 20 mg oral tablet	惊厥	否	是	剂量/剂型文本清洗	
恢复成功	复合扰动	irbesartan	avapro	腹痛	否	是	商品名/别名映射	
恢复成功	复合扰动	prucalopride	motegrity	头痛	否	是	商品名/别名映射	
恢复成功	轻度拼写错误扰动	travoprost	trvoprost	感染/寄生虫感染	否	是	大语言模型拼写纠错	travoprost
恢复成功	轻度拼写错误扰动	levetiracetam	levetiracetam	惊厥	否	是	大语言模型拼写纠错	levetiracetam

4. 讨论

本文结果表明，在 ADR 闭集知识检索任务中，系统先进性不应只由规范化输入下的准确率衡量。对于结构化图谱而言，标准药物名的精确查表本身已经接近上限；更关键的问题是，系统能否在真实用户

输入中识别非规范药物名称，并将其映射回可查询的结构化实体。本文实际解决的是四类更贴近部署的问题：商品名或别名导致的实体失配，剂量/剂型等附加文本导致的标准名匹配失败，轻度错拼导致的图谱入口失效，以及药理类别约束下的集合检索需求。本文通过多关系图谱和分层归一化流程，将系统适用范围从标准药名输入扩展到商品名、剂量文本、复合扰动、拼写错误和药理类别查询。

第一，多关系知识图谱使 ADR 检索从单关系二部图扩展为可表达药物别名和药理类别的结构化知识网络。药物别名节点直接对应真实输入中的商品名和别名，药理类别节点则支持类别级药理查询。与单纯提高大语言模型参数规模相比，这种结构扩展更直接地解决了闭集事实检索中的实体匹配和路径检索问题。

第二，归一化模块是提升真实场景可用性的关键。别名扰动和剂量/剂型文本扰动结果显示，原始精确匹配图检索面对非规范药名时容易出现系统性漏检；一旦加入别名映射和输入清洗，检索性能即可显著恢复。这说明许多真实应用错误并非源于图谱缺少 ADR 关系，而是源于用户输入没有被正确映射到图谱实体。

第三，大语言模型在本文系统中的最佳角色是“检索前规范化”和“检索后表达”，而不是自由生成药物安全事实。拼写纠错实验显示，大语言模型可以有效处理规则难以覆盖的错拼药名；同时，事实标签仍由图谱检索锁定，有助于降低生成式幻觉风险。本文将拼写纠错定位为常规归一化失败后的高鲁棒性回退机制，而非所有查询的默认步骤。

第四，类别级查询使系统从“单个药物是否导致某副作用”扩展到“某类药物中哪些药物与某副作用相关”。本文以 ATC 前缀层级作为药理类别索引，使系统能够在类别节点和药物节点之间进行路径遍历。这类查询更接近临床和药学信息支持中的集合检索需求，也为后续引入频次、严重性和多源证据提供了结构基础。

5. 局限性与未来工作

本文研究虽取得一定进展，但仍存在以下局限。

第一，知识边界与功能定位。当前研究聚焦于 SIDER 派生的闭集知识访问，系统主要功能是可靠提取既有记录，而非发现未编目的新 ADR 关系，亦不能替代严谨的因果推断或药物警戒信号检测。

第二，缺乏独立第三方数据集的外部验证。目前所有评测均属于同源资源内验证，即正例均来自图谱构建的同一知识源。虽然高分结果证明了系统对既有知识的鲁棒访问能力，但其跨来源的外部泛化能力仍需进一步评估。目前的架构已具备良好的模块化特征，未来计划引入完全独立的第三方数据集，如 FDA 药品说明书(Structured Product Labels)或 FAERS 数据库，重点关注系统对未在 SIDER 中收录的罕见 ADR 的检索效果，以及在处理多中心、非规范化临床记录时的稳健性。

第三，模型能效比与纠错模块的深度评估。本实验采用远程调用 Qwen2.5-72B-Instruct 的方案，尚未系统对比 7B/8B 等轻量级端侧模型在同一管线中的表现。同时，虽然 LLM 纠错模块提升了入口恢复率，但针对极短缩写或复杂药名的纠错准确率(Precision)和幻觉风险仍需开展更细致的独立量化评估，以优化成本与延迟的折中。

第四，检索架构的全面性。文本检索增强生成(RAG)的完整同环境复现尚未完全涵盖所有 SOTA 基线。本文重点讨论的是图检索在真实输入变体下的扩展能力，而非主张其在所有 RAG 架构中均具有绝对优势。

第五，复杂实体的建模深度。多成分品牌药(组合制剂)仍是别名归一化的难点。当前的图谱模式在处理“一药多成分”时的解析路径尚不完善，后续需显式建模活性成分层级，并扩展图谱模式以包含剂量、严重程度、发生频率和证据等级。

未来工作将沿两个方向推进：一是扩大知识源，通过跨来源验证提升系统的跨库泛化性能；二是深化知识表达，从单一的关系查询转向支持多维度、高可靠性的药物安全综合问答。

6. 结论

本文提出了一种面向真实世界药物名称变体的增强型图检索增强生成 ADR 检索方法。该方法通过引入药物别名节点和药理类别节点，将原有药物-副作用精确图检索扩展为支持别名归一化、剂量文本清洗、拼写纠错和类别级查询的多关系图检索系统。实验结果显示，增强系统在规范输入下保持高准确率，并在别名扰动、剂量/剂型文本扰动、混合扰动和轻度拼写错误场景中显著提高检索鲁棒性。总体而言，本文为将图检索增强生成从规范化基准推向真实世界 ADR 查询场景提供了一条可复现、可扩展的技术路径。

参考文献

- [1] Nygren, S., Erdogan, O., Avci, P., Daniels, A., Rassool, R., Beheshti, A., *et al.* (2026) RAG-Based Architectures for Drug Side Effect Retrieval Using Compact LLMs. *Scientific Reports*, **16**, Article No. 12754. <https://doi.org/10.1038/s41598-026-41495-2>
- [2] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., *et al.* (2020) Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020*, 6-12 December 2020, 9459-9474.
- [3] Kuhn, M., Letunic, I., Jensen, L.J. and Bork, P. (2016) The SIDER Database of Drugs and Side Effects. *Nucleic Acids Research*, **44**, D1075-D1079. <https://doi.org/10.1093/nar/gkv1075>
- [4] Hogan, A., Blomqvist, E., Cochez, M., D'amato, C., Melo, G.D., Gutierrez, C., *et al.* (2021) Knowledge Graphs. *ACM Computing Surveys*, **54**, Article No. 71. <https://doi.org/10.1145/3447772>
- [5] FDA Adverse Event Reporting System (FAERS) Public Dashboard. <https://fis.fda.gov/sense/sense/app/777e9f4d-0cf8-448e-8068-f564c31baa25>
- [6] 祝赫, 吴凡. 基于 FAERS 的药品不良事件数据挖掘研究现状的文献计量分析[J]. *药物资讯*, 2025, 14(5): 369-378.
- [7] 曾永保, 蒲从福, 徐霞. 226 例老年人群抗精神病药物药品不良反应的回顾分析[J]. *药物化学*, 2025, 13(2): 145-149.
- [8] Kass-Hout, T.A., Xu, Z., Mohebbi, M., Nelsen, H., Baker, A., Levine, J., *et al.* (2015) OpenFDA: An Innovative Platform Providing Access to a Wealth of FDA's Publicly Available Data. *Journal of the American Medical Informatics Association*, **23**, 596-600. <https://doi.org/10.1093/jamia/ocv153>
- [9] ATC/DDD Methodology. <https://www.who.int/tools/atc-ddd-toolkit/methodology>
- [10] Medical Dictionary for Regulatory Activities (MedDRA). <https://www.meddra.org/>
- [11] RxNorm and RxClass APIs. <https://lhncbc.nlm.nih.gov/RxNav/>
- [12] Neo4j Graph Database Platform. <https://neo4j.com/>