

## 知識グラフと GPT を用いた対話的大学情報システムにおける 不正確な入力への対処法の検討

### Handling Inaccurate User Queries to Conversational University Information System using Knowledge Graph and GPT

桑迫 宥登<sup>1</sup> 加藤 恒夫<sup>1</sup> 田村 晃裕<sup>1</sup>  
Yuto Kuwasako Tsuneo Kato Akihiro Tamura

#### 1. はじめに

近年、大規模言語モデル (Large Language Model: LLM) の進展により、自然言語による質問応答や対話型システムの構築が急速に広がっている。LLM は強力な言語生成能力を備えている一方で、事実に基づかない内容を出力するハルシネーションが大きな課題として認識されている。

この問題に対処する手法として、外部知識を参照して応答を生成する Retrieval-Augmented Generation (RAG) が提案されている。RAG は、事前に大規模な知識を学習した LLM に対し、ユーザクエリに応じて関連する情報を検索・補強することで、出力の信頼性と一貫性を向上させる [1][2][3]。これまでの RAG では、主に全文検索やベクトル検索を用いた情報検索が主流であり、比較的単純なクエリに対しては高い精度を発揮してきた。しかし、これらの手法は語彙の一致や意味的類似性に依存しており、意味的には近くとも文脈的な関係性が低い情報を誤って取得することがある [3]。このような誤検索は、特に複雑なクエリや専門的領域において LLM の応答精度を著しく低下させる要因となる。

こうした課題に対し、近年では知識グラフ (Knowledge Graph: KG) を活用した RAG が注目されている。KG はエンティティ間の関係性を明示的に構造化することで、意味的曖昧さを排除し、正確な推論を支援できる。たとえば、KG を活用した RAG では、検索結果の多様性と整合性を同時に確保する手法が提案されている [4][5]。また、実サービスにおいても、LinkedIn で運用されたカスタマー QA システムに知識グラフを組み合わせることで、回答の信頼性とカバー率を向上させた事例が報告されている [6]。

しかし、これらの検索はユーザの入力が正確であることを前提としており、固有名詞の誤記や言い換え、入力省略といった曖昧な表現を含むクエリに対しては、適切な情報を取得できず、誤った応答が生成されるリスクが高い。大学情報システムにおいては、利用者が大学情報に不慣れである場合が多い。従来の検索・生成手法では正確な情報に到達することが困難であるため、不正確な入力にも対応できる情報検索フローが求められる。

本研究では、知識グラフを活用した RAG による大学情報システムを構築する。特に、大学情報に不慣れなユーザが誤記や曖昧な表現で入力を行った場合でも、段階的な検索切り替えにより正確な情報取得が可能となる多段階検索フローを導入することで、LLM の回答の正確性を高める。同志社大学を対象とした情報システムを試作し、LLM 単体・RAG・知識グラフ統合 RAG との比較を通じて、提案手法の有効性を評価する。

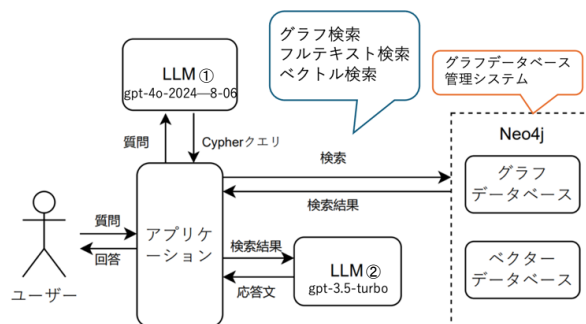


図 1 システム概要 (ベースライン RAG)

## 2. 大学情報システム

### 2.1 システム構成

LLM に外部知識として大学の情報を含む知識グラフを与えることで大学情報 QA システムを構築する。図 1 にシステム構成と応答フローを示す。本システムはユーザに対して入出力を行うアプリケーション、ユーザの自然言語による入力を Cypher クエリに変換する LLM①、大学情報を格納したグラフデータベースの Neo4j、最終的な回答を生成する LLM②によって構成される。

まず、ユーザ入力を LLM① (gpt-4o-2024-08-06) により自然言語から Cypher クエリに変換する。このステップは、Neo4j で情報検索するために必要であり、生成された Cypher クエリによって Neo4j 内を検索し、情報を取得することができる。Neo4j の検索はグラフ検索、全文検索、ベクトル検索の 3 種類で行う。グラフ検索はノード間の関係性をたどり正確な構造情報を取得する。全文検索はノードの説明文プロパティに登録されたテキストを検索する。ベクトル検索はノードのテキスト情報を埋め込みとして表現した embedding プロパティを用いて入力文とのコサイン類似度により検索する。その後、入力文と 3 つの検索結果をまとめて LLM② (gpt-3.5-turbo) に対するプロンプトとして最終的な回答を生成させる。

システムは以下のステップで開発した。まず、著者らが同志社大学に関する質問を 100 問作成し、gpt-3.5-turbo に入力した。その結果、55 問の不正確な出力が確認された。これらの質問に回答できるようにするために、Neo4j で同志社大学知識グラフを構築した。図 2 に構築した知識グラフの概略を示す。キャンパスや学部、学科などのノードを 468 個、[:HAVE], [:ROLE] などのエッジを 658 個作成した。その後、ユーザの入力文を LangChain を用いて、LLM①により Cypher クエリに変換し、Neo4j を検索、検索結果と質

<sup>1</sup> 同志社大学大学院理工学研究所 Graduated school of science and engineering, Doshisha University

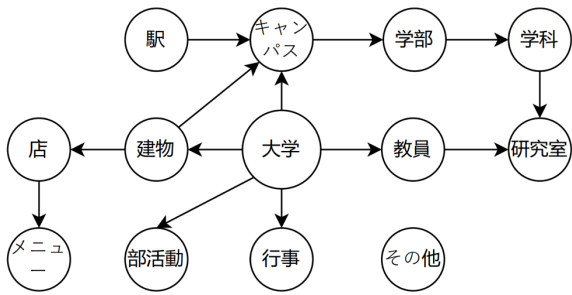


図 2 知識グラフの概略

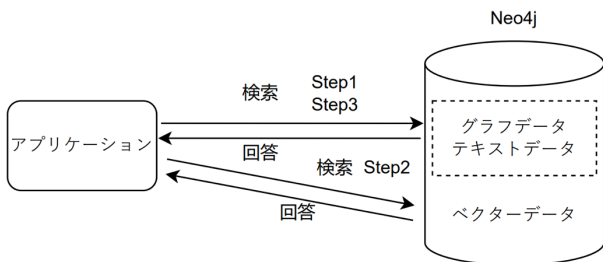


図 3 不正確な入力に対する多段階検索フロー

問を入力として LLM②が最終回答を生成する同志社大学情報システム (ベースライン RAG) を開発した。

## 2.2 不正確な入力に対する多段階検索フロー

ベースラインのシステムでは、3つの検索手法を同時に実行し LLM が回答の統合を行っていた。この場合、不正確な入力を含むクエリではグラフ検索や全文検索は失敗し、ベクトル検索のみが成功する。しかし、類似度は高いが関連性の裏付けがないため正確な回答生成が行われない場合がある。例えば、「パソコン部に入りたい。」という質問に対して、パソコン部が存在しない場合、PC 関連 (情報系の学部学科、PC が利用できる教室) などの情報が出力された。これらは単純に類似性の高い情報であるだけで、質問の本質の部活動とは無関係である。このような場合に、正確な回答を行うためには、質問と回答の関連性の裏付けをすることが必要となる。

そこで、検索の成功・失敗に応じて検索手法を段階的に切り替える多段階検索フローにより、不正確な入力に対しても正確な検索を可能にする。図 3 にユーザの不正確な入力に対する多段階検索フローを示す。多段階検索フローは以下の3つのステップで構成される。

### Step 1: グラフ検索

グラフ検索およびフルテキスト検索を実行し、クエリと一致するノードを取得する。

### Step 2: 類似ノードの取得

Step 1 が失敗した場合、ベクトル検索によりクエリと類似度の高いノードを取得する。

### Step 3: 再検索

Step 2 で取得したノードを起点として、再度グラフ検索およびフルテキスト検索を行い、関連情報を抽出する。以上のフローにより、関連性の不正確な情

報を含む入力に対しても、グラフ検索を正確に実行することができるため知識同士の関係性を考慮した回答の生成が可能となる。

## 3. システム評価

### 3.1 評価実験概要

システムの評価は以下の2点に着目する。まず評価(1)では、RAG 及び、知識グラフを用いた RAG の有効性の検証する。LLM 単独 (gpt-3.5-turbo)、従来 RAG (ベクトル検索、全文検索)、知識グラフを用いた RAG (ベースライン) を比較し RAG の有効性を評価する。

次に評価(2)では、不正確なユーザ入力に対する有効性の検証を行う。提案手法の多段階検索フローによる不正確な入力時の有効性をベースラインのシステムと多段階検索フローのシステムを比較し評価する。

### 3.2 評価データセット

本学の様々な学部所属する大学生 10 人から同志社大学に関する質問を集めた。質問は、学部や研究室、部活動、施設、学食など多様なトピックを含んでおり、入力文の総数は 133 個となった。これを評価(1)ベンチマーク用のテストセットとする。評価(1)では、133 個の質問に対して回答に漏れがなく、正確な回答ができていないか評価する。例を表 1 に示す。例の出力は、正確な情報かつ漏れがないため正しい回答である。この情報が 1 つでも欠けていた場合は正しくない情報として評価する。

また、評価(1)で使用したテストセットを基に文章の一部 (主に固有名詞) を不正確なものに変更したものを 100 個作成した。評価(1)で使用した 133 個の質問の中には、固有名詞などを持たない入力があったためこれらは省いている。例えば、「キャンパスはどこにありますか?」のような質問は不正確な入力になる可能性がないためこのような入力は省いた。例を表 2 に示す。これを評価(2)のテストセットとする。評価(2)では、作成した 100 個の質問に対して、知識グラフを用いた RAG における改良した多段階検索フローが有効であるかベースラインと比較し評価する。

表 1 正確な解答例

入力	理工学部にある学科を教えてください。
出力	同志社大学理工学部には以下のような学科があります: 1. **インテリジェント情報工学科** (中略) 2. **情報システムデザイン学科** (中略) 3. **電気工学科**  (以下省略)

表 2 不正確な入力例

正確な入力	京田辺キャンパスにある学部を教えてください。
不正確な入力	田辺キャンパスにある学部を教えてください。

表3 評価(1)の応答種別の割合（133問）

ラベル	正しい回答	正しくない回答	回答不可（DBに情報なし）	計
LLM単独（gpt3.5-turbo）	22%	78%		100%
RAG（全文検索，ベクトル検索）	57%	19%	24%	100%
ベースラインRAG（グラフ検索，全文検索，ベクトル検索）	65%	11%	24%	100%

表4 評価(2)の応答種別の割合（100問）

ラベル	正しい回答	正しくない回答	回答不可（DBに情報なし）	計
ベースラインRAG（グラフ検索，全文検索，ベクトル検索）	54%	22%	24%	100%
多段階検索フローRAG（グラフ検索，全文検索，ベクトル検索）	66%	10%	24%	100%

表5 各応答例の比較

入力	京田辺キャンパスのパン屋はどこにありますか？
LLM単独	同志社大学京田辺キャンパス内には、焼きたてパンを取り扱うパン屋があります。以下のとおりです。 ・ <b>ベーカリーカフェ Doshisha</b> (以下省略)
RAG（全文検索，ベクトル検索）	同志社大学京田辺キャンパス内および周辺には、焼きたてパンを楽しめるお店がいくつかあります。以下に代表的な店舗をご紹介します。 1. <b>ベーカリーカフェ Johermi</b> （ジョアミ） (中略) 2. <b>良心館ベーカリーカフェ</b> (以下省略)
ベースラインRAG	京田辺キャンパスのパン屋は、ローム記念館の1階にある「 <b>Johermi</b> （ジョアミ）」です。 (以下省略)

表6 ベースラインRAGと多段階検索フローRAGの応答例の比較

入力	同志社大学の情報工学部について教えてください。
ベースラインRAG	情報工学部についての情報を Neo4j から検索しましたが、結果は得られませんでした。したがって、情報工学部が存在するかどうかについては確認できませんでした。 (以下省略)
多段階検索フローRAG	同志社大学に <b>情報工学部は存在しません</b> が、以下の学部学科が確認できました。 **文化情報学部** (中略) **インテリジェント情報工学科** (中略) **情報システムデザイン学科** (以下省略)

### 3.3 評価(1)の結果

表3に評価(1)の結果として、133問の質問に対して各応答種別の割合を示す。正しい回答の割合は、LLM単独（gpt3.5-turbo）の22%の回答精度に対して、RAG（全文検索，ベクトル検索）を適用することで57%と大幅に性能が向上したことがわかる。さらに、知識グラフを適用することでRAG（全文検索，ベクトル検索）よりもベースラインRAG（グラフ検索，全文検索，ベクトル検索）が高い性能を発揮している。以上から、RAG，知識グラフを用いたRAGによってLLMの回答精度を高めることができ、有効性が示された。特に、学部と学科，キャンパスと施設のように、明確な階層的关系を持つ情報に対しては、KGを用いたRAGの有効性が顕著に現れた。一方、DBに情報があるにも関わらず正しくない回答となったものも確認された。これらの多くは略称や独自に使っている名称が含まれた質問であり、結果的に正しくない回答となった。例えば、「文化情報学部」を「文情」と略したり、「インテリジェント情報工学科」を「イン情」とすることで正確な検索ができず、正しくない回答の生成となった。また、システムの性能とは無関係に、対象とする情報がデータベースに含まれていなければ、正確な応答を返すことは不可能となるため、データベースの拡張が今後の改善における重要な課題であることが再認識された。

表5にLLM単独，RAG（全文検索，ベクトル検索），ベースラインRAG（グラフ検索，全文検索，ベクトル検索）の応答例を示す。また、表5の下線部は各応答のハルシネーション部分を示している。LLM単独ではハルシネーションが生じており事実とは異なる情報を生成している。このようにLLM単独の応答ではハルシネーションが多く含まれ、信頼性に欠ける出力が多く見られた。RAGでは、一定程度改善できたものの、正確な応答とはならなかった。表5ではパン屋を正確に出力できているものの、別キャンパスのパン屋が混じっており、正確な回答とは言えない。一方でベースラインRAGでは、KGによってキャンパスとパン屋との関係性の裏付けができ、入力に対する必要な情報のみを出力としているため正確な応答である。これにより、知識グラフを用いることで関係性を考慮した回答に有効であることが分かる。

### 3.4 評価(2)の結果

表4に評価(2)の結果として、100問の不正確な入力に対するベースライン RAG と多段階検索フローRAG の応答の割合を示す。評価①と同様、データベースにない情報を含む質問が24%あり、これらは全て回答不可能となる。正しい回答の割合は、ベースライン RAG (グラフ検索、ベクトル検索、全文検索) の精度が54%に対して、改良後の多段階検索フローによる RAG の精度が66%になり、多段階検索フローが有効であることがわかる。グラフ検索失敗時に意味的に近いノードを探索し、それを起点として構造的に再探索を行うという検索アプローチが有効であることを示した。一方で、不正確な入力にもかかわらずベースライン RAG の54%の正しい回答は、回答に関係性を必要としない質問であり、ベースラインのベクトル検索で十分な質問が多くあったためであると考えられる。

表6にベースライン RAG と多段階検索フローRAG の応答例を示す。例のように「同志社大学の情報工学部について教えてください。」という不正確な入力に対して、ベースライン RAG では情報工学部がデータベース内に存在しないため、回答不可という結果になった。これは、本来回答となりうる情報がデータベースに存在しているにもかかわらず適切な検索が行われなかったからである。しかし、多段階検索フローRAG では「情報工学部」という語に含まれる意味的な特徴から関連するノード(インテリジェント情報工学科、情報システムデザイン学科、文化情報学部など)を適切に抽出し、さらにそれらのノードを基点に構造的検索を行うことで、実在する関連情報を提示することに成功した。不正確な入力を訂正しつつ、関連する情報をデータベースから取得できていることが分かる。以上から提案手法の多段階検索フローが QA システムにおける不正確な入力に対して有効であることが示された。以上の結果から、提案手法である多段階検索フローRAG は、大学情報に不慣れなユーザによる不正確なクエリに対しても正確な情報取得を実現できることが確認された。今後、より多様な曖昧表現への対応や、クエリ前処理段階での誤字補正・類義語展開などを取り入れることで、さらなる精度向上が期待される。

### 4. 終わりに

本研究では、知識グラフと LLM を統合した対話型の同志社大学 QA システムの構築を行なった。知識グラフを活用することで、大学の複雑な情報構造を正確に表し、従来の RAG では困難であった構造的関係を考慮した応答生成が可能であることを確認できた。加えて、大学情報に不慣れなユーザが利用することを想定した、不正確な入力に対する多段階検索手法においても、有効性を確認できた。これは、実運用における多様なユーザが利用することを前提としたシステム設計において成果であると考えられる。今後の実用化に向けては、情報の拡張、インタフェースの改善、音声入力への対応などの開発を進める予定である。

#### 謝辞

助言と議論をいただいた、CDM Inteact の Graham Wilcock 氏、産業技術総合研究所の Kristiina Jokinen 氏に感謝いたします。

#### 参考文献

- [1] Y. Gao et al., "Retrieval-Augmented Generation for Large Language Models: A Survey," arXiv preprint arXiv:2312.10997, 2023.
- [2] S. Gupta et al., "A Comprehensive Survey of Retrieval-Augmented Generation (RAG): Evolution, Current Landscape and Future Directions," arXiv preprint arXiv:2410.12837, 2024.
- [3] W. Fan et al., "A Survey on RAG Meeting LLMs: Towards Retrieval-Augmented Large Language Models," arXiv preprint arXiv:2405.06211, 2024.
- [4] X. Zhu et al., "Knowledge Graph-Guided Retrieval Augmented Generation," arXiv preprint arXiv:2502.06864, 2025.
- [5] L. Bahr et al., "Knowledge Graph Enhanced Retrieval-Augmented Generation for Failure Mode and Effects Analysis," arXiv preprint arXiv:2406.18114, 2024.
- [6] Z. Xu et al., "Retrieval-Augmented Generation with Knowledge Graph for Customer Service Question Answering," Proc. SIGIR '24, pp. 2905–2909, 2024.