

知識グラフと RAG を活用した 規制化学物質含有推定 AI の高度化

山本隼輔¹ 綱川隆司¹ 西田昌史¹

¹ 静岡大学大学院総合科学技術研究科情報学専攻

{yamamoto.shunsuke.20@, tuna@inf., nishida@inf.}@shizuoka.ac.jp

概要

環境規制が厳格化する中、第三者部品を含む製品の化学物質調査は難しく、製品構成情報の機密性から外部 API の利用が困難である。本研究はローカル LLM に検索拡張生成 (RAG) とナレッジグラフ (KG) を組み合わせ、製品に含まれる物質を推定する手法を提案する。BM25 検索¹⁾と KG フィルタで誤検出を抑え、各構成要素の有効性を検証した結果、KG・LLM・RAG の統合により高い適合率と F1 スコアを達成した。禁止物質に限らず将来規制される物質も予測し、プライバシーを保持したまま調査を支援する。

1 はじめに

環境リスクを低減するために製品に含まれる物質に関する世界的な規制が強化され続ける中、これらの物質の含有の有無を調査することは、特に第三者の製造業者の部品が含まれる場合にはますます困難になっている。さらに、製品の構成部品に関する情報は機密であることが多く、製品情報を外部に送信することなく調査を行うことが困難である。

規制物質を予測する必要性は、いくつかの課題から生じている。国によって異なる物質に対して規制が施行されており、規制物質リストは毎年更新され、新たな物質が追加されている。古い製品に使用されている物質が現在規制されているかどうかは不明確な場合が多い。製造が第三者企業に委託されている場合、詳細な情報の開示が不確実であったり、企業秘密であることを理由に開示されなかったりすることがある。化学分析により確定的な結果が得られるものの、コストとリソースがかかる。

これらの課題に対処するため、我々は、検索拡張生成 (RAG) とナレッジグラフ (KG) を統合したロー

カル大規模言語モデル (LLM) を用いて、製品内の物質を予測する手法を提案する。本研究の主な目的は、部品および材料データをドキュメントとして使用し、これらの推論の精度を向上させることである。KG を使用しないローカル LLM、KG のみの手法、その他の LLM と KG の組み合わせなど、他の手法と比較することで、提案手法の有効性を評価した。

我々の貢献は、KG と独立した BM25 検索による RAG と、製品に接続されない物質を除去する事後 KG フィルタを組み合わせ、プライバシーを保つローカル LLM による予測が可能であることを実証した点にある。KG は生成後のフィルタリングにのみ用いられ、LLM 単独・KG 単独・フィルタなし RAG より適合率と F1 を大幅に向上しつつ競争力のある再現率を維持することを示した。

2 関連研究

我々の以前の研究 [1] では深層学習で高精度を実現したが、大規模学習が必要で予測理由の説明を提供できなかった。LLM と RAG は追加学習なしに予測でき、推論根拠も示せる。

Da Silveira ら [2] や Sepehri ら [3] は KG と LLM で有害物質評価を効率化したが、一般的データベース対象で企業の機密情報は扱わない。Hartung [4] は 87% 精度を達成したが説明可能性には課題が残る。

Lewis ら [5] や Gao ら [6] は知識集約型タスクで RAG の有効性を示したが、機密データに対しては未対応である。Zeng ら [7] はプライバシー保護 RAG を提案したが KG 統合は扱っていない。

Chattoraj ら [8]、Hernandez ら [9]、Hoang ら [10] は KG ベースのコンプライアンス推論を提案したが、LLM と RAG の統合は未検討である。Hu ら [11] や Peng ら [12] の Graph RAG は検索時にグラフ活用するが、本研究は事後フィルタで高精度を実現する。

1) BM25 (Best Matching 25) は、文書の関連性をスコアリングする情報検索アルゴリズム。

本研究は、ローカル LLM、RAG、KG フィルタを組み合わせ、プライバシー保持と高精度を両立し、LLM 単独・KG 単独・フィルタなし RAG などの比較手法よりも適合率と F1 を向上させる。

3 データセット

本研究で使用したデータセットは、ヤマハ発動機株式会社から提供された chemSHERPA²⁾ 準拠の製品含有化学物質情報である。製品情報には、製品名、製造会社、部品、物質、各部品と物質の質量が含まれる 40,430 製品で構成されている。本研究では、製品名、部品名、物質データを用いて物質の有無を予測した。予測タスクは、物質の量を推定することなく、その有無を判定することに限定した。

4 提案手法

4.1 手法概要

本研究の主な目的は、ローカル LLM、RAG、KG を活用して、製品に含まれるすべての物質を予測することである。類似した部品や材料を持つ製品は、共通の物質を持つ可能性が高いという仮説を立てた。この仮説に基づき、RAG を介して検索された関連製品、部品、材料に関する情報で LLM への入力を強化し、KG を用いて LLM の出力から無関係な物質を除外する手法を提案する (図 1 参照)。

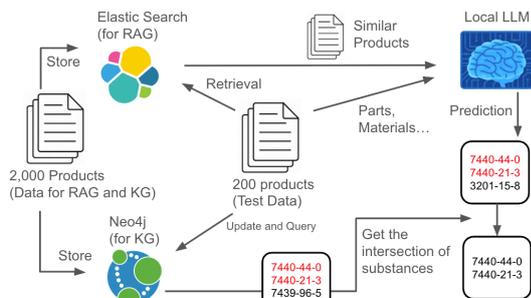


図 1 提案手法のフローチャート

まず、含有物質予測を行うテストデータの製品からその製品が含む部品や材料の情報を抽出する。この情報は、Elasticsearch^[13] の BM25 クエリを構築するために使用され、重複する部品や材料を持つ類似製品を表す関連ドキュメントを検索する。LLM による予測を行う前に、部品、材料、物質間の関係を含むテストデータの製品の情報がグラフデータベース Neo4j^[14] に挿入され、関連物質の KG ベースの

2) chemSHERPA は、国際的な化学物質含有情報フォーマット標準。

検索が可能になる。

KG では、製品、部品、材料、物質がノードとして表現され、HAS PART、HAS MATERIAL、HAS SUBSTANCE などの関係がエッジとして表現される。この構造により、関連情報の柔軟かつ効率的な検索が可能になる。具体的には、製品の物質を予測する際、システムはグラフを探索して、これらのエッジを介して製品ノードに接続されているすべての物質を検索する。検索プロセスは製品ノードから始まり、HAS PART および HAS MATERIAL エッジをたどって部品および材料ノードに到達し、HAS SUBSTANCE 関係を通じて接続されているすべての物質を収集する。このアプローチにより、システムは製品の構造情報に基づいて関連物質を効率的に検索できる (図 2 参照)。



図 2 製品物質予測のための KG 構造。オレンジ、緑、ピンク、青のノードはそれぞれ製品、部品、材料、物質を表す。エッジは HAS PART、HAS MATERIAL、HAS SUBSTANCE などの関係を示す。

検索されたドキュメントと物質は、予測を生成するためのコンテキストとして LLM に提供される。予測の信頼性を確保するため、LLM の出力は、KG から検索された物質との共通部分をとることで後処理される。全体的なプロセスは以下の通りである。

- 1. Local LLM + KG + RAG (提案手法):** ローカル LLM、Neo4j KG、Elasticsearch ベースの RAG を組み合わせる。LLM は検索された製品ドキュメントに基づいて予測を生成する。出力は KG の結果を使用してフィルタリングされ、無関係な物質が除去される。
- 2. Local LLM Only (Baseline 1):** 検索されたドキュメントや KG に頼らず、製品情報のみを入力として LLM を使用する。
- 3. KG Only (Baseline 2):** Neo4j を使用して、テスト製品の関連物質を直接検索する。LLM や RAG は関与しない。
- 4. Elasticsearch Only (Baseline 3):** BM25 検索によ

る RAG のみを使用する。検索されたドキュメントに含まれるすべての物質を予測物質とし、LLM や KG は適用しない。

5. **Local LLM + KG (Baseline 4):** RAG による文書検索を組み込まず、LLM と KG の知識を入力として使用する。
6. **Local LLM + RAG (Baseline 5):** RAG コンポーネント (BM25 検索) と LLM のみを使用する。KG は除外され、検索された製品ドキュメントのみに依存して予測を行う。

提案手法は Haystack フレームワーク [15] を用いて実装されている。Neo4j [14] は製品、部品、材料、物質に関する構造化情報を保存し、Elasticsearch は RAG のドキュメントストアとして機能する。ローカル LLM として、Llama-3.1-Swallow-8B-Instruct-v0.5 [16] を使用した。

BM25 は、共有部品や材料に基づいて関連ドキュメントを検索するためのシンプルかつ効果的な手法として採用された。そのシンプルさにもかかわらず、BM25 は、複雑な検索手法を必要とせず、LLM の予測を強化するドキュメントの検索において強力なパフォーマンスを示した。

予測結果に同一の物質が繰り返し現れる現象など、LLM の一般的な問題に対処するため、LLM 出力から重複を除去する後処理を適用した。最後に、KG によるフィルタリングにより、テスト製品に関連する物質のみが残るようになり、予測精度と信頼性が向上する。

4.2 データ準備

予測性能を評価するために、知識源としてのドキュメントとして 2,000 サンプル、テストデータとして 200 サンプルを準備した。その内訳は、部品が 1,721、材料が 66、物質が 573 である。まず、製品情報がデータセットから抽出され、Elasticsearch と Neo4j の両方に保存された。その後、ローカル LLM にテストデータをクエリとして提供し、製品に含まれる物質を予測させた。

データ準備プロセスの詳細を以下に説明する。ドキュメント、クエリ、質問の形式の詳細については、付録 A を参照されたい。

4.3 Neo4j による物質検索

KG から物質を検索するために、テスト製品の構造に基づいて Neo4j にクエリを実行する。具体的に

は、部品や材料を介して製品に接続されている物質を抽出し、構造的に関連する物質のみが LLM の出力フィルタリングの対象となるようにする。

ナレッジグラフは、ヤマハ発動機株式会社から提供された chemSHERPA 準拠の製品情報を Neo4j にマッピングすることによって構築された。これらの内部的に管理されている情報は、各企業の従業員によって作成およびレビューされている。ただし、軽微なノイズ (例: 一貫性のない、または曖昧な部品名) や情報の欠落が含まれている場合がある。

5 実験

本節では、実験設定を示し、製品内の物質を予測するための提案手法を使用して得られた結果の詳細な分析を提供する。

5.1 評価手法

LLM の出力から予測物質リスト (例: [8934-34-3, 2342-23-1]³⁾) を取得し、各テスト製品の正解ラベルと比較する。

提案手法の有効性を評価し、他のベースライン手法と比較するために、評価指標として適合率 (Precision)、再現率 (Recall)、F1 スコアを使用した。適合率は、予測されたすべての物質のうち正しく予測された物質の割合を測定し、予測の正確さを示す。再現率は、実際のすべての物質のうち正しく予測された物質の割合を測定し、関連するすべての物質を特定するモデルの能力を反映する。F1 スコアは適合率と再現率の調和平均である。

評価は一括で実施し、Elasticsearch Only を含む全手法の予測性能を同一テーブルで比較した。各手法について同一実験を 3 回繰り返し、その平均値を結果として報告した。

環境設定の詳細は付録 C に記載されている。

5.2 LLM による説明付き予測

LLM に物質予測だけでなく、各予測の根拠となる理由も同時に生成させる実験を行った。これにより、予測結果の解釈可能性を向上させることができる。代表的な説明例を付録 B に示す。

3) CAS 番号 (Chemical Abstracts Service 番号) は、化学物質を一意に識別する国際的な登録番号である。

表1 各手法の性能比較 (2,000 件の学習データと 200 件のテストサンプルを使用)

Method	Micro Average			Macro Average		
	Precision	Recall	F1	Precision	Recall	F1
Local LLM	0.020	0.060	0.030	0.043	0.072	0.044
KG Only	0.084	0.962	0.155	0.150	0.902	0.229
Elasticsearch Only	0.794	0.905	0.846	0.812	0.868	0.821
Local LLM + KG	0.338	0.060	0.102	0.211	0.072	0.092
Local LLM + RAG	0.784	0.741	0.761	0.791	0.749	0.747
Local LLM + KG + RAG	0.902	0.746	0.817	0.832	0.756	0.774

5.3 結果と考察

表1は、各手法の詳細な性能（マイクロおよびマクロの適合率、再現率、F1 スコアを含む）をまとめたものである。

実験結果より、異なる手法間での再現率と適合率のトレードオフが明らかになった。KG のみのアプローチは最高の再現率 (**0.962**) を達成する一方で、適合率は極めて低い結果 (0.084) となった。これは、フィルタリングなしで製品の部品や材料に関連するすべての物質を検索するため、多くの偽陽性が生じるためである。そのため、この手法は網羅性には優れるものの、正確な予測に必要な精度を欠いている。一方、提案手法 (LLM + KG + RAG) は適合率と再現率のバランスを取ることで、高い適合率 (**0.902**) と F1 スコア (**0.817**) を達成した。

すなわち、KG は網羅性には優れるものの、無関係な物質を除去する推論能力がないため、高い再現率の一方で低い適合率となる。これに対し、LLM と RAG を統合することで、文脈的な推論と絞り込み検索が可能となり、偽陽性が減少して適合率が大幅に向上する。さらに、テスト製品の KG に存在しない物質を除外する後処理により、予測が構造化された知識に基づいた信頼性の高いものになる。

表1から注目すべき点として、Elasticsearch Only はマイクロ F1 スコア 0.846 を達成し、Local LLM + RAG (0.761) および Local LLM + KG + RAG (0.817) の提案手法をも上回る F1 スコアを取得している。しかし、Elasticsearch 単体では予測理由が得られず、適合率は 0.794 に留まり、提案手法の 0.902 よりも大幅に低い。つまり、Elasticsearch 単体は網羅的な検索性には優れるが、予測根拠を提示できず、且つ誤検出が多い。この結果は、単純な関連製品の抽出が目的なら Elasticsearch 検索システムで十分だが、予測理由の説明と結果の正確性が必要な場合は、LLM と KG を統合した提案手法が有効であることを示唆

している。

ローカル実装により、データの機密性が確保される。これは製品構成に機密情報が含まれる業界にとって極めて重要である。LLM をローカルにデプロイし、構造化された検索メカニズムを活用することで、我々のアプローチは高精度でプライバシーを保護した推論を実現し、実用的価値を示している。

一方、提案手法にはいくつかの課題がある。LLM が物質を繰り返し生成することがあり、現在は重複を除去する後処理で対処しているが、根本的には LLM の生成過程に起因するものと考えられる。

付録 B に示したように、LLM は材料名や材料クラスを根拠に物質の含有理由を具体的に説明できる一方で、一部の物質については同じ説明が繰り返される傾向が見られた。説明付き予測は予測結果の信頼性を向上させ、特に製造業のコンプライアンス判定で重要だが、説明の重複削減は今後の課題である。

6 おわりに

本研究は、ローカル LLM に KG と RAG を統合して製品内物質を予測する手法を示した。KG 単独は高再現率だが低適合率、ローカル LLM 単独はリソース不足で不十分だったのに対し、統合手法は構造化知識と推論を併用して F1 を大幅に改善し、限られた計算資源下でもプライバシーに配慮した有効性を確認した。

今後の課題として以下のことが挙げられる。

- **スケーラビリティ:** 40,430 製品までデータを拡張し、検索・予測パイプラインを最適化する。
- **説明可能性:** 物質 ID とともに推論根拠を生成し、受容性を高める。
- **出力品質:** 重複生成を抑制するプロンプト・生成制御を検討し、精度への影響を評価する。

これらにより、製品コンプライアンスと持続可能性に向けた実用性をさらに高める。

謝辞

本研究はヤマハ発動機株式会社の支援を受けました。データセットおよび技術的な助言をいただいた坂下大祐氏、櫻井雄一氏に感謝します。

参考文献

- [1] 山本隼輔, 鈴木佳那子, 坂下大祐, 櫻井雄一, 綱川隆司. 深層学習による規制化学物質の含有推定. In **Proceedings of the 87th National Convention of IPSJ**, 2025.
- [2] Marcos Da Silveira, Louis Deladiennee, Kheira Acem, and Oona Freudenthal. Combining knowledge graphs and llms for hazardous chemical information management and reuse. In **Proceedings of IEEE International Conference on Bioinformatics and Biomedicine (BIBM)**, pp. 6766–6773, Lisbon, Portugal, 2024. IEEE.
- [3] Sara Sepehri, Anja Heymans, Dinja De Win, Jan Maushagen, Audrey Sanctorum, Christophe Debruyne, Robim M Rodrigues, Joery De Kock, Vera Rogiers, Olga De Troyer, and Tamara Vanhaecke. Toxin graph: A semantic knowledge graph for animal-free toxicological risk assessment. **Comput. Toxicol.**, 2025.
- [4] Thomas Hartung. Artificial intelligence as the new frontier in chemical risk assessment. **Frontiers in Artificial Intelligence**, Vol. 6, p. 1269932, 2023.
- [5] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Kuttler, Mike Lewis, Wen tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive NLP tasks. In **Proceedings of NeurIPS 2020**, 2020.
- [6] Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, Meng Wang, and Haofen Wang. Retrieval-augmented generation for large language models: A survey. **arXiv preprint**, 2023. Preprint <https://arxiv.org/abs/2312.10997>.
- [7] Shenglai Zeng, Jiankun Zhang, Pengfei He, Jie Ren, Tianqi Zheng, Hanqing Lu, Han Xu, Hui Liu, Yue Xing, and Jiliang Tang. Mitigating the privacy issues in retrieval-augmented generation (rag) via pure synthetic data. **arXiv preprint**, 2024. Preprint <https://arxiv.org/abs/2406.14773>.
- [8] Rittwik Chatteraj and Amit P. Joshi. Medreg-kg: Knowledge graph for streamlining medical device regulatory compliance. In **Proceedings of IEEE International Conference on Big Data**, pp. 3382–3390, Washington, DC, USA, 2024. IEEE.
- [9] Julio Hernandez, Delaram Golpayegani, and Dave Lewis. An open knowledge graph-based approach for mapping concepts and requirements between the eu ai act and international standards. **AI Ethics**, 2025. <https://doi.org/10.1007/s43681-025-00708-6>.
- [10] Van Thuy Hoang, Tien-Bach-Thanh Do, Jinho Seo, Seung Charlie Kim, Luong Vuong Nguyen, Duong Nguyen Minh Huy, Hyeon-Ju Jeon, and O-Joun Lee. Halal or not: Knowledge graph completion for predicting cultural appropriateness of daily products. In **IEEE Access**, pp. 15158–15167, Access 13, 2025. IEEE. <https://arxiv.org/abs/2501.05768>.
- [11] Yuntong Hu, Zhihan Lei, Zheng Zhang, Bo Pan, Chen Ling, and Liang Zhao. Graph-rag: Graph retrieval-augmented generation for multi-hop reasoning. **Proceedings of AACL**, 2025.
- [12] Boci Peng, Yun Zhu, Yongchao Liu, Xiaohe Bo, Haizhou Shi, Chuntao Hong, Yan Zhang, and Siliang Tang. Graph retrieval-augmented generation: A survey, 2025. Preprint <https://arxiv.org/abs/2408.08921>.
- [13] Elasticsearch B.V. Elasticsearch. **Software**, 2025. <https://www.elastic.co/jp/elasticsearch/>.
- [14] Inc. Neo4j. neo4j. **Software**, 2025. <https://neo4j.com/>.
- [15] deepset. Haystack. **Software**, 2025. <https://haystack.deepset.ai/>.
- [16] AI@Meta. Llama 3 model card. **Model Card**, 2024. https://github.com/meta-llama/llama3/blob/main/MODEL_CARD.md.

A データフォーマット

A.1 サンプルドキュメント

サンプルドキュメントには、製品 ID、会社名、部品情報、材料情報、物質情報が含まれる。

製品 ID: SWITCH

開発された都道府県: 静岡県

開発された市区町村: 浜松市

開発した会社名: ヤマハ発動機株式会社

部品情報は以下の通りです。

部品名: スイッチボディー

部品名: 端子ネジ

部品名: 接点

材料情報は以下の通りです。

材料名: Copper alloys

材料 ID: T523

材料クラス ID: Y-102

物質情報は以下の通りです。

物質名: Copper (Cu)

物質 ID: 7440-50-8

A.2 サンプル質問

LLM への入力プロンプトは、製品情報と部品・材料情報を含む。説明付き回答の場合は、回答形式としてカンマ区切りの代わりに「各予測の根拠を生成してください。」という指示が追加される。

この製品 (製品 ID: SWITCH) に含まれる可能性の高い物質を教えてください。

ここで言う物質とは「carbon」「silicon」などで、材料のことではありません。

また、物質を答えるときは、「7439-95-4」のような ID で回答してください。

回答の形式は、カンマ区切りで「xxxx-xx-x,xxxx-xx-x,xxxx-xx-x」のように回答してください。

同じ物質は 1 回だけ回答してください。

製品情報は以下の通りです。

製品 ID: SWITCH

開発された都道府県は静岡県

開発された市区町村は浜松市

開発した会社名はヤマハ発動機株式会社

部品情報は以下の通りです。

部品名: スイッチボディー

部品名: 端子ネジ

部品名: 接点

材料情報は以下の通りです。

材料名: Copper alloys

材料 ID: T523

材料クラス ID: Y-102

B LLM による説明付き予測の例

LLM に物質予測と各予測の根拠を同時に生成させる実験を行った。表 2 に代表的な例を示す。

表 2 LLM による説明付き予測の例 (電子部品)

物質 (CAS 番号)	LLM が生成した説明
7440-50-8 (銅)	導体、ニッケルめっき、金めっきの材料として使用されるため。
7440-02-0 (ニッケル)	導体、ニッケルめっきの材料として使用されるため。
7440-50-8 (銅)	材料名に「Copper」と記載されているため。
7440-21-3 (シリコン)	PET (P519) の材料として使用される可能性が高い。
7440-21-3 (シリコン)	filled Thermoplastics (N499) の材料として使用される可能性が高い。

LLM は材料名や材料クラスを根拠に、物質が含まれる理由を説明できる。これは RAG で検索された類似製品情報を LLM が正しく理解していることを示し、予測結果の信頼性を向上させる。説明付き予測は製造業のコンプライアンス判定で特に重要である。

C 実装環境

提案手法の実装環境を表 3 にまとめる。これには、使用されたハードウェア、ソフトウェア、フレームワークの詳細が含まれる。

表 3 実装環境

Category	Details
Operating System	Ubuntu 20.04.6 LTS
RAM	128 GB
CPU	Intel(R) Core(TM) i9-10940X CPU @ 3.30GHz
GPU	2 × NVIDIA RTX A6000 (48 GB VRAM each)