

地域資料における災害データのナレッジグラフを用いた表現と その利用方法の提案

Proposal for Representation and Use of Knowledge Graphs for Disaster Data in Regional Materials

堀本隆誠^{1*}, 岡田真¹, 森直樹¹

Ryusei HORIMOTO^{1*}, Makoto OKADA¹, Naoki MORI¹

1 大阪公立大学

Osaka Metropolitan University

〒599-8531 大阪府堺市中区学園町1-1

E-mail: sc24847p@st.omu.ac.jp

*連絡先著者 Corresponding Author

近年の人工知能技術の急速な発展の中で、人間の知識をグラフ構造で表現するナレッジグラフが注目されており、異なるデータソースからの情報の統合やそれらの間の関係性を明確にするために有効である。本研究では、地域資料の保存・継承における災害データの管理手法として、ナレッジグラフを用いた手法を提案する。ナレッジグラフを用いることで、過去の種々の災害の記録における事象のデータとそれらの相互関係を統一して記述可能であり、これにより複数の情報の統合と再利用性の向上による知識の継承を容易にすると予想される。さらに、個々の災害データの相互関連性の可視化により、将来的な災害リスク評価や災害対応の改善にも寄与することが期待される。本稿においては、実際のデータに対してナレッジグラフを構築し、考察することでその可能性を図る。

In recent years, artificial intelligence technology has rapidly developed, with knowledge graphs, which represent human knowledge in a graph structure, gaining significant attention. These graphs are utilized as foundational technologies in various fields and tools. Knowledge graphs effectively integrate information from diverse data sources and clarify the relationships between them. This study proposes a method for managing disaster data in the preservation and inheritance of regional materials using knowledge graphs. By representing various past disaster events and their interrelations through a unified structure, knowledge graphs are expected to facilitate the integration of multiple sources of information, thereby enhancing

data reuse and knowledge transfer. Furthermore, by visualizing the interrelationships between individual disaster data, this approach is anticipated to contribute to future disaster risk assessments and improve disaster response strategies.

キーワード：災害データ, ナレッジグラフ, 大規模言語モデル, GPT

Keywords : disaster data, Knowledge Graph, Large Language Model, GPT

1 はじめに

近年、人工知能技術は急速な発展を遂げている。その中で大規模言語モデル (Large Language Model: LLM) および人間の知識をグラフ構造で表現するナレッジグラフ[1] が注目を集めている。特に、OpenAI 社が発表した LLM をベースとした GPT-4o[2] はさまざまな自然言語処理タスクで優れた成果を残している。また、ナレッジグラフはさまざまな知識とそのつながりをグラフ構造として表現するデータ構造であり、多種多様な情報とそのつながりを体系的に表現できるという利点や、自然言語文のような非構造データを情報として扱えるという利点がある。また、ナレッジグラフを用いることで複数のデータに含まれる知識や情報を可視化して集約しやすくなる。災害情報のような時系列や地域で変化が生じるデータについても、それらからの知識抽出や集約において有効だと考えられる。

本研究では、GPT-4o を利用して地域資料における災害データからナレッジグラフを生成する手法と利用方法について提案する。

2 要素技術

2.1 ナレッジグラフ

ナレッジグラフ[1] とは、さまざまな知

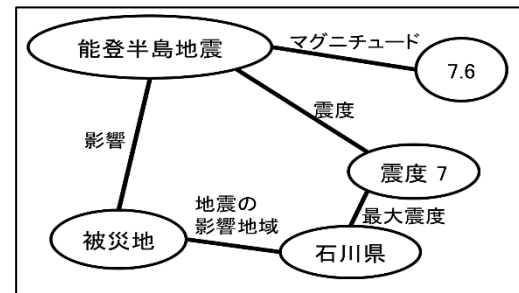


図1 ナレッジグラフの例

識を体系的に連結し、その関係をグラフ構造で表した知識ネットワークのことである。図 1 にナレッジグラフの例を示す。ナレッジグラフは、エンティティとそのエンティティ間の関係性を表現するリレーションから構成されており、図 1 のようにエンティティをノード、リレーションをエッジとするグラフとして表される。また、ナレッジグラフを表す表現方法として (エンティティ 1, リレーション, エンティティ 2) という 3 つ組構造 (トリプル) の集合で表すこともできる。ナレッジグラフはさまざまな種類のデータを統合的にグラフとして扱うため、一度に多面的な情報を得ることができる。

2.2 内閣府による災害データ

内閣府による災害データ[3] は日本国内で発生した自然災害に関する情報を網羅的に提供しているデータである。このデータには、災害の発生日時、場所、規模、被害状況、対応措置などが含まれ、各種報告書や統計資料が PDF 形式でまとめられ

ている．本研究では 2024 年 1 月 1 日に発生した能登半島地震の報告書のうち，1 月 2 日および 8 月 21 日時点のものを使用する．

3 提案手法

ナレッジグラフはさまざまな知識とそのつながりをグラフ構造として表現するデータ構造であり，異なるデータセットの相互関係を明確にすることができる．本研究では，過去の種々の災害データセットを GPT-4o へ入力して，その結果としてナレッジグラフの各エンティティとリレーションをトリプル形式で抽出し，それらを統合することによりナレッジグラフを構築する手法を提案する．図 2 に提案手法のモデル概略図を示す．なお，PDF 形式の災害データは GPT-4o への入力のためにテキストデータに変換する．

4 実験

提案手法の有効性を確認するために，実験 1 として災害データセットからのトリプル抽出，実験 2 として抽出されたトリプルからのナレッジグラフ構築という 2 種類の実験をした．以下では 2 種類の実験について説明する．

実験 1 では，内閣府による災害データにおける 2024 年 1 月 1 日に発生した能登半島地震の報告書のうち，1 月 2 日および 8 月 21 日時点のものを GPT-4o への入力とし，プロンプトで指示したトリプルを抽出した．プロンプトの主な設定は以下のようになっている．

- 災害報告書からナレッジグラフを作成
- 重要なエンティティの抜き出し
- エンティティ同士の関係性の推測
- 抽出

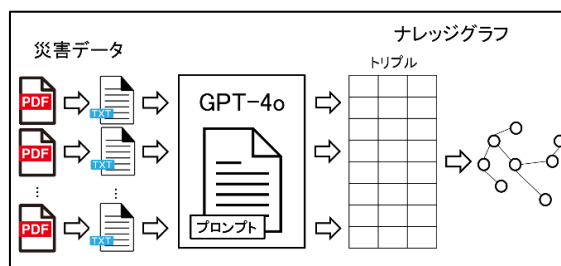


図2 モデル概略図

したトリプルの重要度（1~5 の整数）の設定．なお，本研究では GPT-4o の入力トークンの関係上，災害報告書の最初の 10 ページのみを入力としている．

実験 2 では，内閣府による災害データから実験 1 で抽出したトリプルを統合し，ナレッジグラフを可視化した．なお，実験 1 で設定した重要度はグラフにおけるノードの大きさに対応させる．

5 実験結果と考察

5.1 実験結果

表 1, 2 に実験 1 で 1 月 2 日および 8 月 21 日時点の報告書から抽出したトリプルの一例をそれぞれ示す．表 1, 2 より，“石川県-最大震度-震度7”のように地震における重要なトリプルが共通して抽出できていることがわかった．一方で，“令和6年能登半島地震-地震の深さ-震源の深さ”や“石川県-復旧の完了-復旧完了”のようにエンティティと関係性が重複するトリプルが抽出されていることもわかった．

図 3 に実験 2 で可視化したナレッジグラフの一部を示す．図 3 より，地震の発生源である“石川県”と地震の関係性が豊富に存在するようにナレッジグラフが作成されていることがわかった．

表1 1 月 2 日時点の報告書のトリプル

エンティティ1	関係性	エンティティ2	重要度
石川県	最大震度	震度7	5
令和6年能登半島地震	地震の深さ	震源の深さ	4
被害内容	地震による二次災害	火災	3
被災地	地震の影響地域	兵庫県	2

表2 8 月 21 日時点の報告書のトリプル

エンティティ1	関係性	エンティティ2	重要度
石川県	最大震度	震度7	5
石川県	地震による被害	被害内容	5
石川県	物流の影響	物流遅延	3
石川県	復旧の完了	復旧完了	4

5.2 考察

GPT-4o はテキストデータから比較的適切なトリプルを抽出できることがわかった。一方、うまく関係性をとれないトリプルも抽出されており、このようなトリプルが抽出されないようにプロンプトを修正する必要があると考えられる。

本研究では報告書作成時刻の異なるデータを扱ったが、時系列による違いは可視化したナレッジグラフには反映されなかった。今後の課題として時系列を考慮できるナレッジグラフの可視化手法を検討する必要がある。

6 まとめと今後の課題

本研究では作成時点の異なる災害報告書からトリプルを抽出し、ナレッジグラフを可視化した。今後の課題として、より適切なトリプル取得のためのプロンプト調整および時系列によるナレッジグラフ

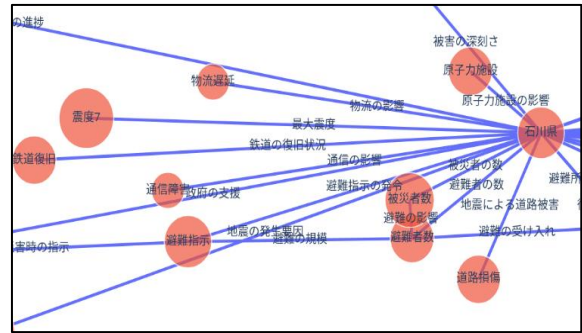


図 3 災害データを統合したナレッジグラフ
の変化の可視化が挙げられる。

7 データ利用可能性宣言

本研究で作成および分析に用いたデータセットは内閣府ホームページの<https://www.bousai.go.jp/updates/index.html> で入手可能である。

謝辞

本研究は、日本学術振興会科学研究補助金基盤研究(C)(課題番号23K11252)の補助を得て行われたものである。

参考文献

- [1] 川村隆浩;江上周作;田村光太郎;外園康智;鵜飼孝典;小柳佑介;西野文人;岡嶋成司;村上勝彦;高松邦彦;杉浦あおい;白松俊;張翔宇;古崎晃司:「第1回ナレッジグラフ推論チャレンジ2018開催報告—説明性のある人工知能システムを目指して—」, 人工知能, 34 巻, 3号, pp. 396-412, 2019.
- [2] OpenAI: “Gpt-4 Technical Report”, <https://arxiv.org/pdf/2303.08774>, (2024 年 10 月 18 日参照)
- [3] 内閣府ホームページ:災害情報, <https://www.bousai.go.jp/updates/index.html>, (2024 年 10 月 18 日参照)