

【背景・目的】 ChatGPT の公開以来、大規模言語モデル(LLM)の能力を発揮させる Prompt Engineering が注目を浴びている。例えばナレッジグラフ(KG)分野においては、prompt による LLM からの情報抽出が有効である¹。諸知識の体系的リンクにより生成されるナレッジグラフ(KG)では、トリプル(テキスト3要素からなるセット)を用いて情報を連結させる。本研究では、汎用性の高いこのトリプルを活用した KG 分析法の開発を目指した。

【結果・議論】 分析例として、3年間の有価証券報告書を用いた結果を示す。図1は、3年間の抽出データより得られた KG の抜粋である。まず初めに、各段落の題名を取得した。次に、それを用いて Few-Shot によりトリプル(題名, 名詞, 説明)を取得した。得られたグラフは、本 KG 分析法の情報俯瞰における有用性に加えて、ハルシネーションを防ぐ RAG(Retrieval-Augmented Generation)への有益性を示唆している。

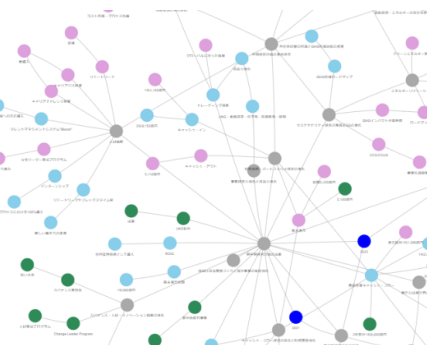


図 1, 有価証券報告書のグラフ

[1] Xu, D.et al. (2023). Large Language Models for Generative Information Extraction: A Survey. *arXiv preprint arXiv:2312.17617*.