Knowledge graph representation of financial relations using SLM

Gustav Brøndumbo-Therkildsen[1], Hagai Ofer[1]

1 DTU Institute for Electronics Fotonics, Technical University of Denmark;

Introduction

Analyzing financial relationships is crucial for auditing, concerning anomaly detection, and decision-making. Deep learning models excel at processing unstructured data and extracting complex patterns but often lacks transparency. This limitation is critical in financial auditing, where decisions must be explainable and traceable.
To address this, we developed a pipeline that uses fine-tuned language models (T5, BART, Flan-T5) to extract relationships from financial data and represent them in a directed knowledge graph. This approach enhances interpretability by visually mapping key nodes, such as transactions, accounts, and anomalies.

Furthermore, graph transformation, enabling:

- Conversion of raw SLM outputs into structured graph components.
- Interaction with the graph using natural language queries.
- Refinement of graph structures for clarity and usability.

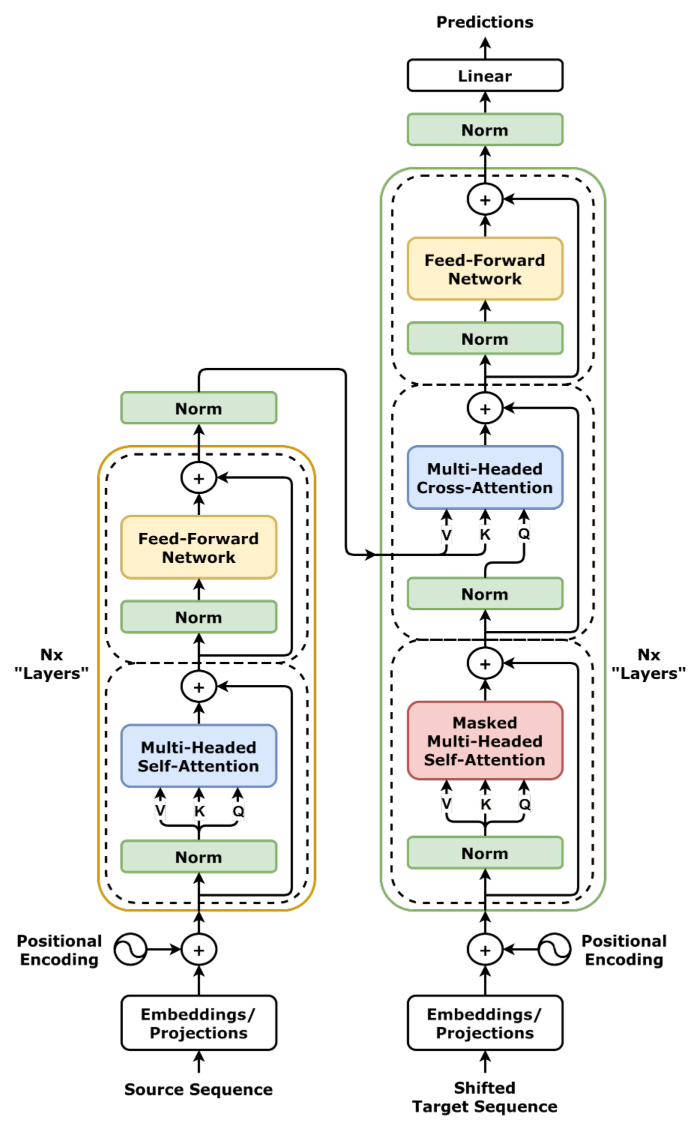
This combination of SLMs and other graph analysis methods creates a transparent, user-friendly tool for uncovering relationships and anomalies in financial data.

Key points

- We fine-tuned 3 different SLMs on a repeatedly generated **structured dataset**.
- Synthetic structured data is generated to ensure the fine-tuning dataset maintains the required format and contains realistic financial relationships.
- We train the models to extract the core sentences.
- We then use **NER** to post-process entity relation triplets, which are then graphed.

SLM and Transformers

Transformers are advanced machine learning models that excel in processing sequential data by leveraging self-attention to capture contextual relationships. In the proposed pipeline, a pre-trained transformer is fine-tuned to summarize input text and extract key sentences, effectively condensing raw financial data into focused, context-rich segments. This step not only refines the data but also generates structured outputs that are easier for downstream components to process, facilitating the identification of entities, relationships, and actionable insights. The fine-tuning process tailors the model to the specific requirements of financial and auditing tasks, enhancing accuracy and relevance for domain-specific applications.



Extracting name entity relations

Initial implementation was just the name entity relation, which resulted in a chaotic mess.

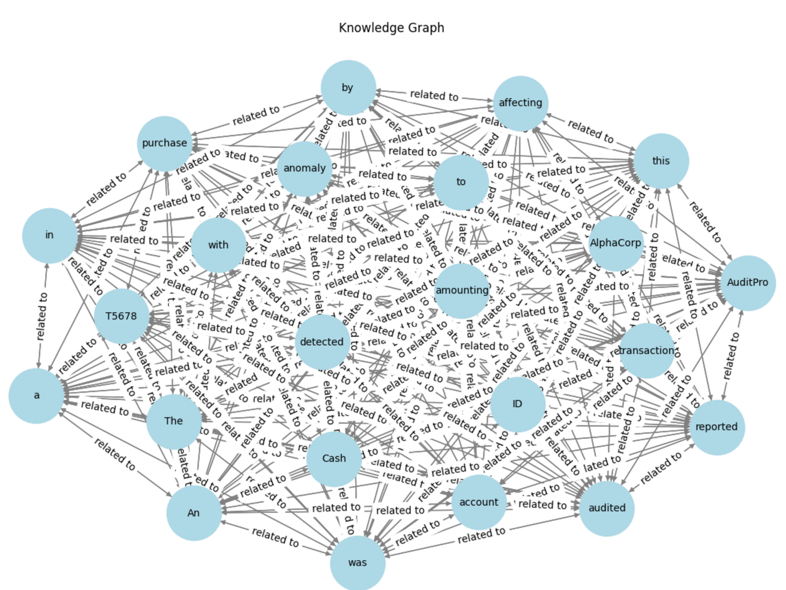


Figure 1: The result of pure name entity relation and no key entity extraction by the SLM

Desired output

The desired output would involve the key elements, such that the user would be able to quickly ascertain where to focus their efforts. One of the ways this would be done is through anomalies, that are found in extension of already implemented software that takes vouchers and investigates them line by line.

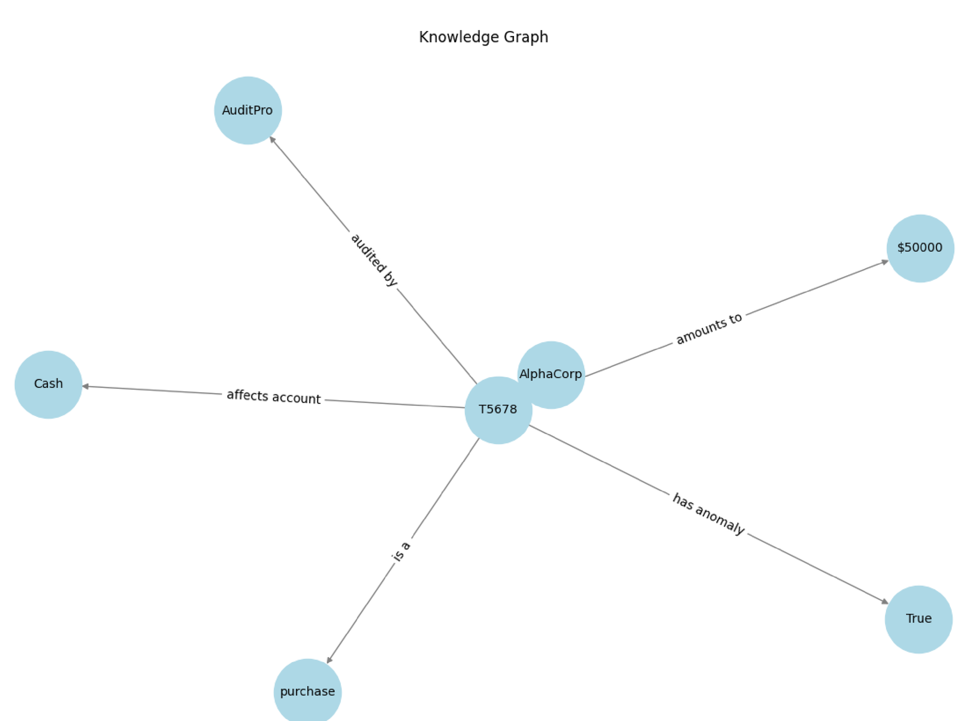


Figure 2: Results of something by another thing.

Pipeline Overview

The proposed pipeline extracts, represents, and visualizes financial relationships from unstructured text. Following these steps:

- **Summarization:** A transformer model condenses unstructured text into key sentences, retaining only the most relevant information.
- **Entity and Relation Extraction:** SpaCy, combined with custom financial patterns, identifies financial entities such as transactions and accounts, and maps their relationships (e.g., "audited by," "affects accounts").
- **Relationship Structuring:** Extracted relationships are formatted into triplets (*Entity A - Relationship - Entity B*) for further processing.
- **Graph Visualization:** Relationships are visualized as a directed knowledge graph using NetworkX, enhancing interpretability.
- **Evaluation and Fine-Tuning:** The pipeline supports evaluation against ground truth datasets and can be fine-tuned for domain-specific tasks, ensuring adaptability and robustness.

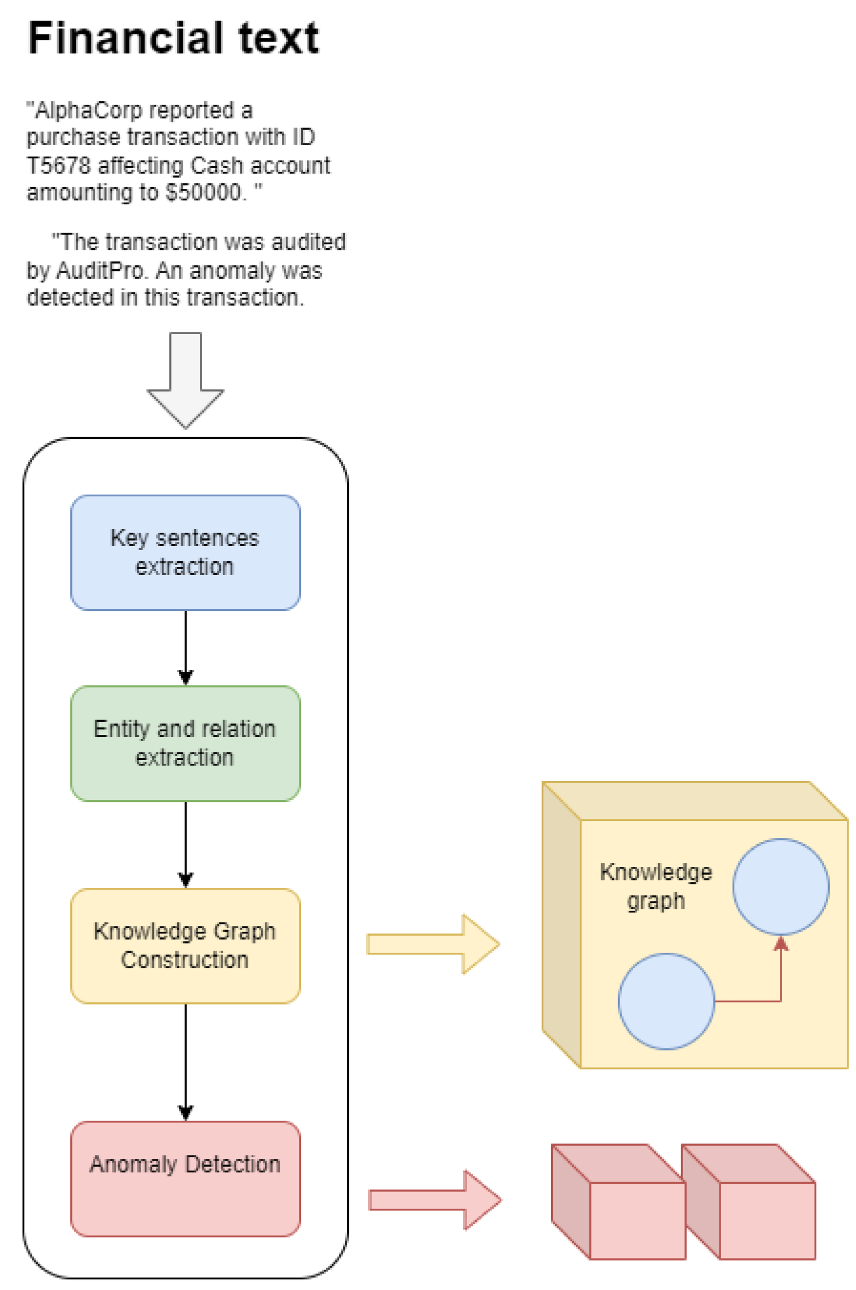


Figure 3: Model architecture representation.

Pipeline performance

The three models t5-small, google/flan-t5-small, and facebook/bart-base achieve precision scores between 0.63 and 0.91, recall scores between 0.47 and 0.53, and F1-scores between 0.54 and 0.66. Google/flan-t5-small attains the highest precision (0.91) and overall F1-score (0.66), while t5-small achieves the highest recall (0.53).

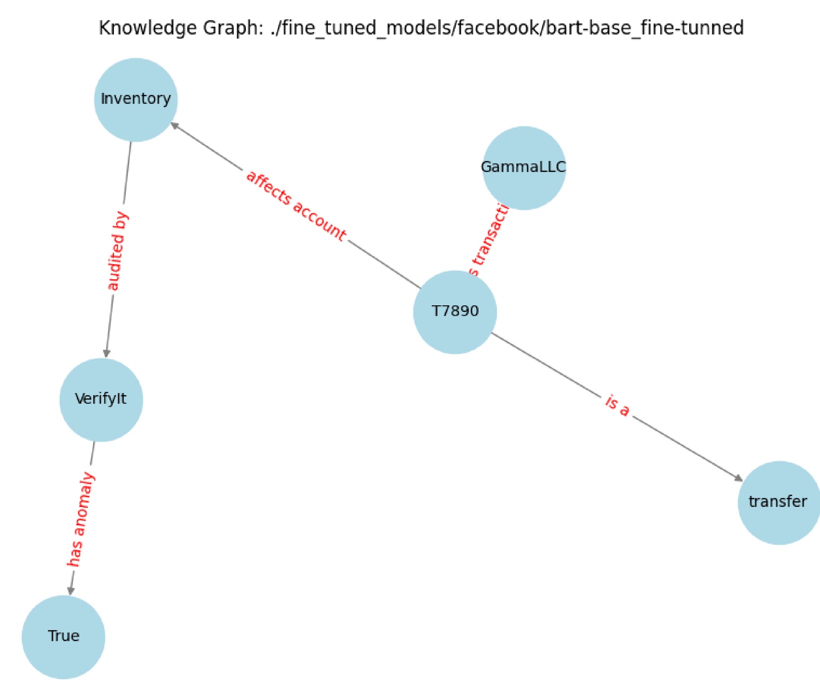
Table 1: Final metrics of the pipeline with the different models.

Model	Precision	Recall	F1-Score
t5-small	0.69	0.53	0.59
google/flan-t5-small	0.91	0.52	0.66
facebook/bart-base	0.63	0.47	0.54

Results

The generated knowledge graphs illustrate how large language models can extract and represent financial relationships among various nodes such as assets, transactions, and organizations from unstructured auditing data. For instance, they identify that certain transactions (e.g., T5678, T7890) 'affect accounts' or are 'audited by' particular entities, while also linking concepts like 'Cash', 'Inventory' or 'VerifyIt' and classifying events as 'transfer', 'purchase', or containing 'anomaly'. These extracted relations, though still lacking full transparency and interpretability, highlight the potential of deep learning methods to pinpoint nodes of interest and uncover hidden connections in financial datasets.

By applying LLM graph transformations, we aim to refine these initial structures into clearer, more usable graphs. This approach not only fosters better understanding of complex auditing relationships but also bridges the gap between raw financial data and actionable insights. As a result, these transformed graphs can serve as a more transparent, navigable resource, ultimately improving our ability to trace transactions, identify anomalies, and support decision-making in auditing and financial analysis.



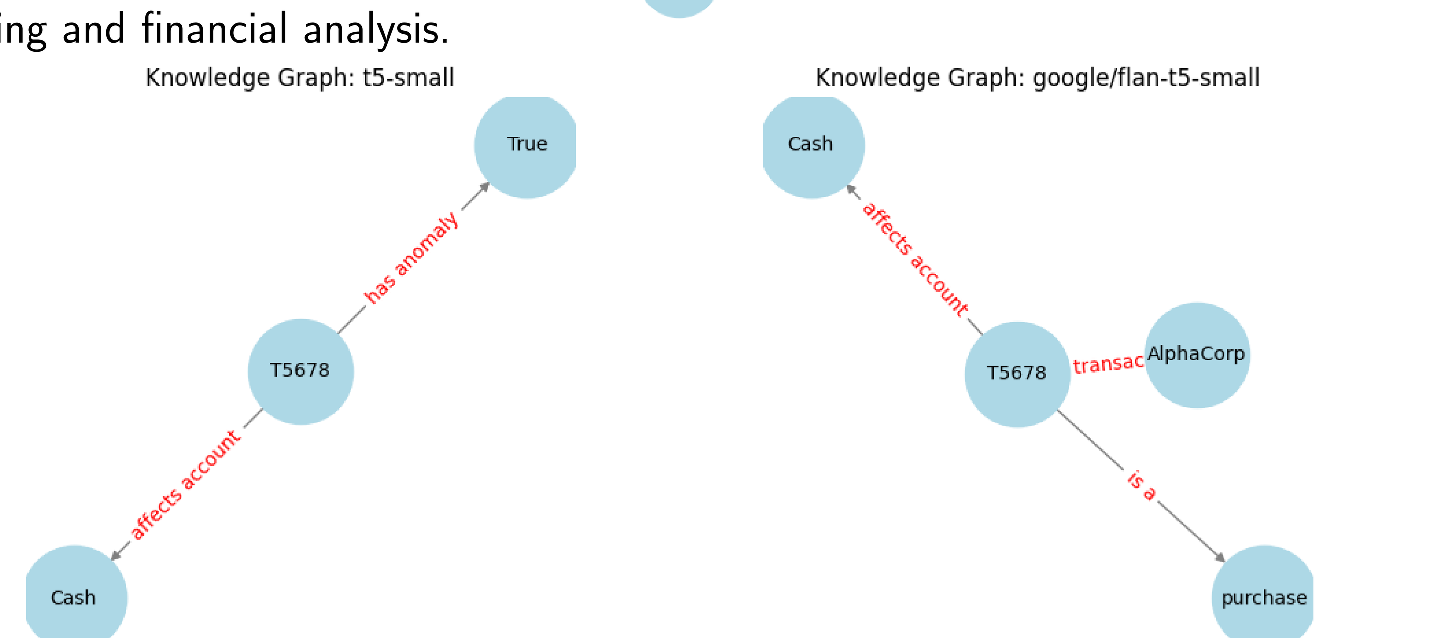


Figure 4: Caption

References

[1] F. AI. Building a custom ner model with spacy: A step-by-step guide, 2023. URL https://blog.futuresmart.ai/building-a-custom-ner-model-with-spacy-a-step-by-step-guide. Accessed: 2024-12-09.

[2] R. L. Amrani et al. From unstructured text to causal knowledge graphs: A transformer-based approach. *arXiv*, 2023. URL https://arxiv.org/abs/2202.11768. Accessed: 2024-12-09.

[3] S. Documentation. Training custom ner models in spacy, 2023. URL https://www.machinelearningplus.com/nlp/training-custom-ner-model-in-spacy/. Accessed: 2024-12-09.

[4] Y. Gao et al. Graph neural network-based entity extraction and relationship reasoning in financial texts. *arXiv*, 2024. URL https://arxiv.org/pdf/2411.15195. Accessed: 2024-12-09.

s233347@dtu.dk, s233249@dtu.dk