Comparing Foundations: Insights into the Construction of Financial Causal Knowledge Graphs with and without Ontology

Author: Ziwei Xu, Ryutaro Ichise

Presenter: Ziwei XU

Date: May 29th 2024

AIST
NATIONAL INSTITUTE OF ADVANCED INDUSTRIAL SCIENCE AND TECHNOLOGY (AIST)

東京工業大学
Tokyo Institute of Technology

Content

1. Introduction

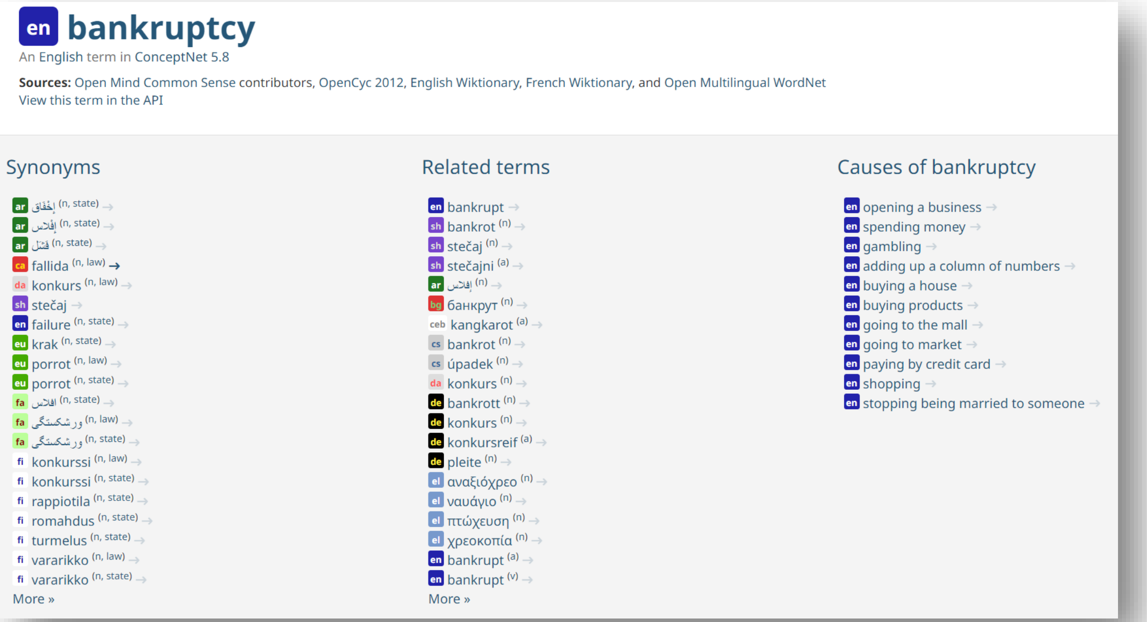
Motivation:

- Causality analysis offer valuable insights for risk mitigation, investment decisions, and portfolio optimization.
- It invokes the generation of mono-relation causality knowledge graphs (mono-rel KGs).
- However, mono-rel KG lacks hierarchical connection to present knowledge in structural manner

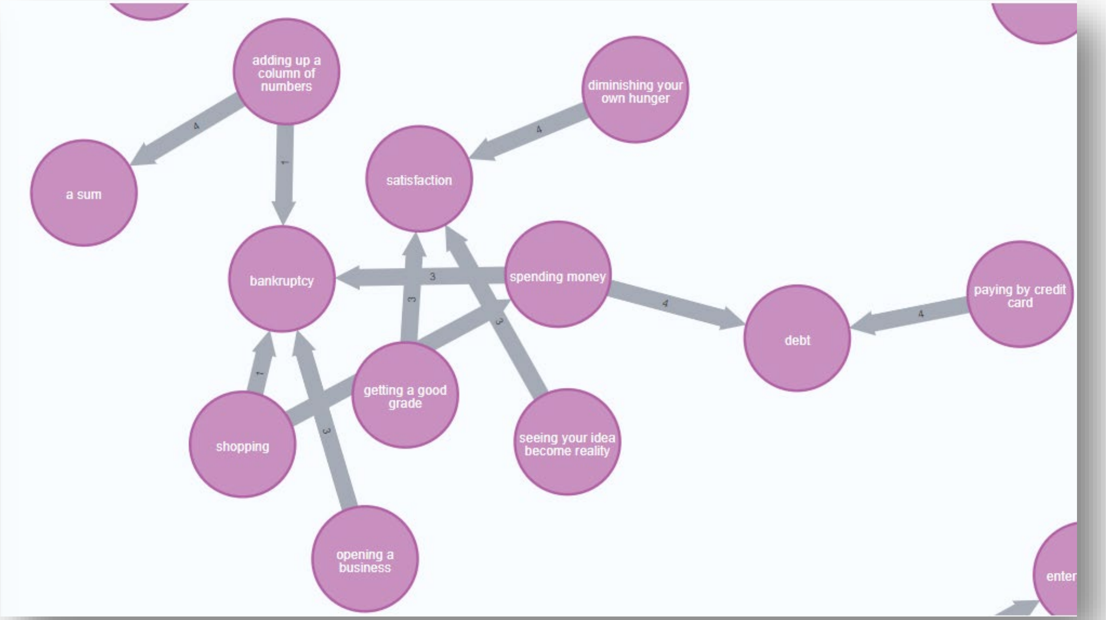
Contributions:

- We exemplify a mono-relation KG through FinCaKG framework (Financial Causality Knowledge Graph)
- We propose our methodology to integrate an expert ontology (FIBO) into FinCaKG --- FinCaKG-Onto
- By comparing these two resulting knowledge graphs, we found, ontology integration expands the scope of nodes and links and can also assist to inference to get further knowledge.

2. Background



Commonsense KG: ConceptNet [1]



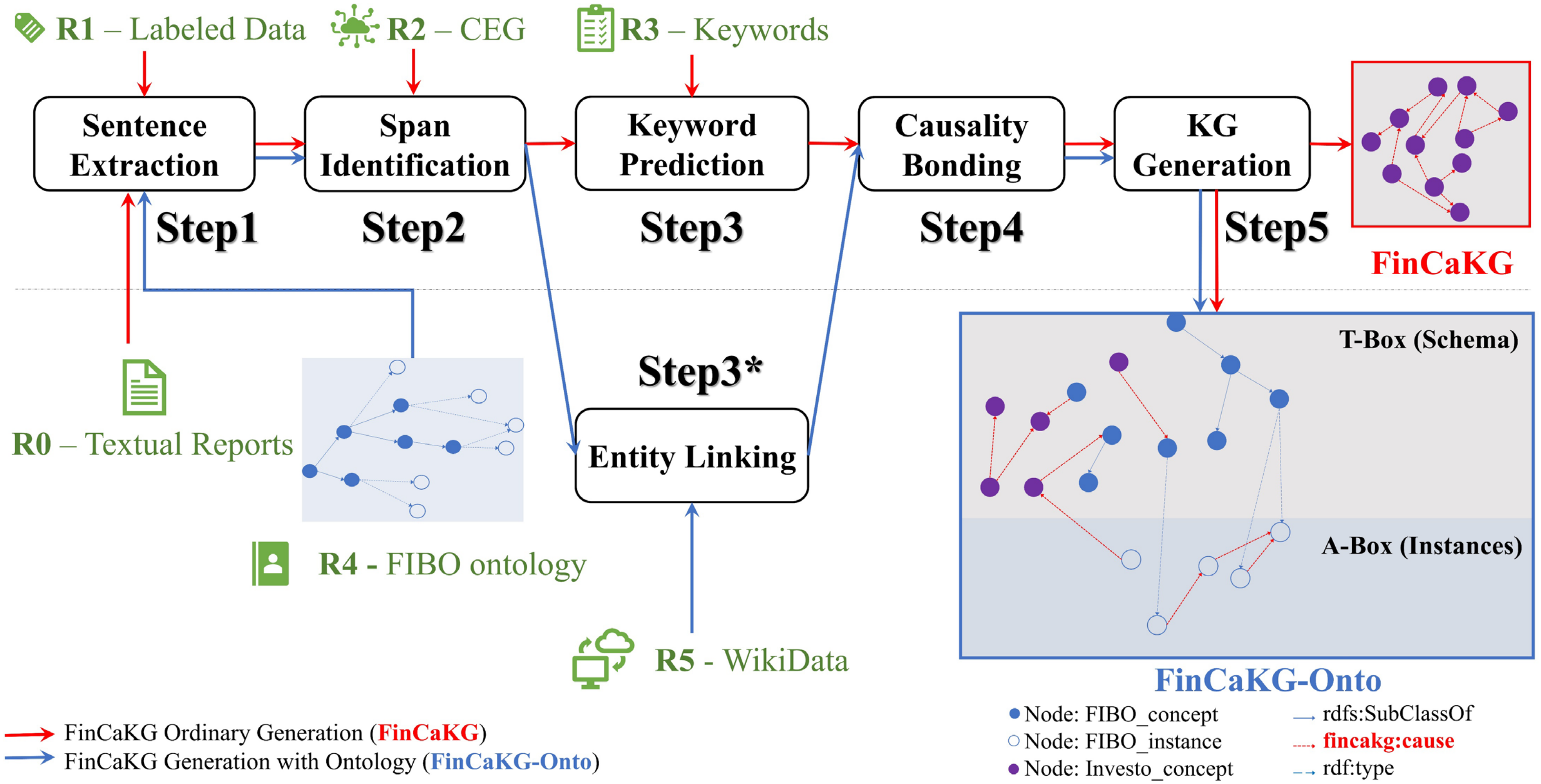
Mono-rel KG: Cause@ConceptNet

Cause@ConceptNet

- Causality here is scattered, common-sense and action-based
- Causality is contextless, thus prone to be doubtable

[1] R. Speer, et al.“Conceptnet 5.5: An open multilingual graph of general knowledge,” AAAI 2017

3. FinCaKG Construction with & without Ontology - Steps



Description of constructive steps:

- **Step1**: extract causal sentences
- **Step2**: identify cause-effect textual spans from causal Sentences
- **Step3**: predict financial keywords of cause-effect spans
- **Step3***: locate the financial keywords in textual spans and link mentions to WikiData entities
- **Step4**: formulate the keyword causality dependency from cause-effect spans
- **Step5**: generate and visualize FinCaKG

3. FinCaKG Construction with & without Ontology - Steps

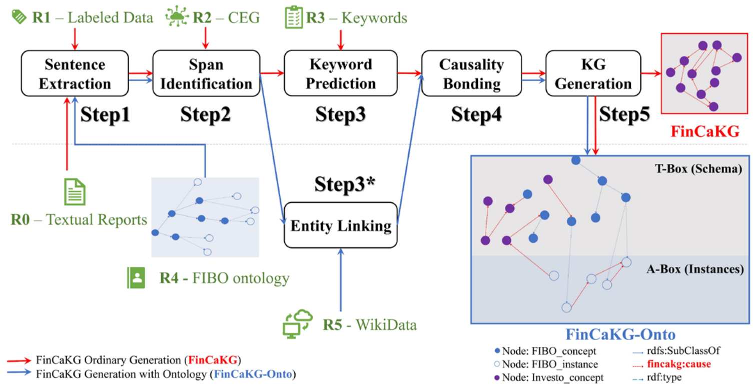
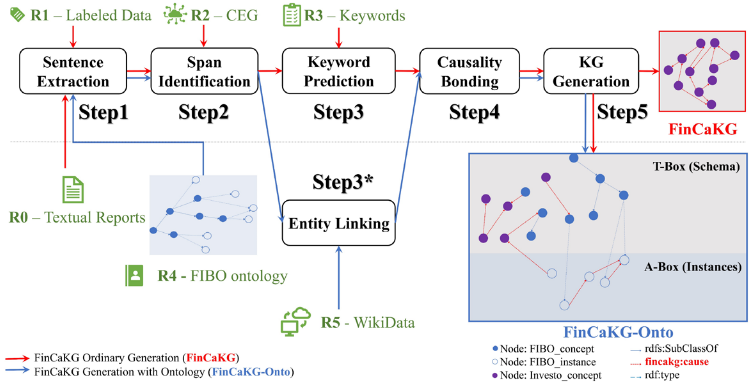


Table 1: The examples that showcase the dataflows within the FinCaKG Generation Framework.

Steps	Input	Output
Step1	**Passage**: Sent1, Sent2... Sent5[The increase in provision for bad debts during fiscal 2019 primarily related to $2.7 million of client OTC derivative account deficits in the Commercial segment, and $1.4 million in the Institutional segment, partially offset by client recoveries in the Commercial segment.] ... Sent10 ... SentN	**Sent5** is a causal sentence
Step2	**Sent5**: The increase in provision... in the Commercial segment.	**Cause Span**: $2.7 million of client OTC derivative account deficits in the Commercial segment, and $1.4 million in the Institutional segment, partially offset by client recoveries in the Commercial segment. **Effect Span**: The increase in provision for bad debts during fiscal 2019
Step3	**Cause Span**: $2.7 million ... derivative ... **Effect Span**: The increase ... bad debts ...	mention "bad debts" is financial keyword "**Bad debt**"
Step3*		mention "derivative" is a class in FIBO ontology "**Derivative (finance)**"
Step4	**outputs** of Step2, Step3 and Step3*	**a causal pair**: "Bad debt" — "Derivative (finance)"
Step5	**all cause pairs** w/wo ontology	**FinCaKG; FinCaKG-Onto**

3. FinCaKG Construction WITH Ontology

- Datasets



Textual Reports (R0)

The financial report in S.E.C of the top 3000 company in the last 5 years.

Labeled Data (R1)

The causal labels for sentences in financial reports.

Cause Effect Graph - CEG (R2)

Cause Effect Graph contains 89.1 million causal pairs.

Financial Keywords (R3)

We consider Investopdia's vocabulary are financial Keywords.

FIBO Ontology (R4)

Financial Business Ontology (FIBO) defined 1100 concepts and 32,134 instances by experts.

WikiData(R5)

We regard it as an entity dictionary in our entity linking tasks.

3. FinCaKG Construction WITH Ontology
- Experiment Settings

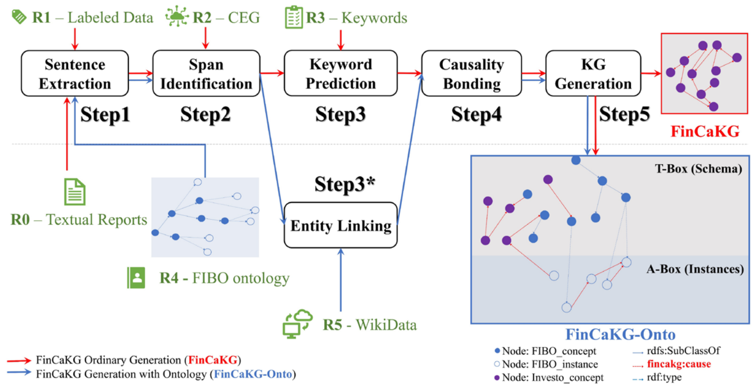


Table 2: The parameters of models in each step.

Steps	Labels	Models	ep.	bt.	lr.	opt.
Step1	0, 1	BERT	10	8	2e-5	epsilon =1e-8
Step2	B-C, I-C, B-E, I-E, O	ilab-FinCau [Xu+22]	5	4	5e-5	
Step3	B-F, I-F, O	XLM-Roberta	5	8	2e-5	
Step3*	[START_ENT], [END_ENT]	GENRE; mGENRE	-	-	-	-

Step3* (Entity Linking Task)

Target: link the potential mentions to financial keywords, and then map vocabulary to Wikidata entities

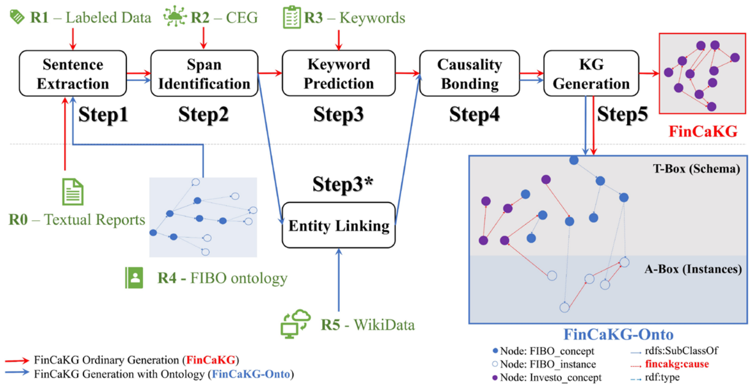
Solutions:

- Method 1: string matching to find keywords and then use GENRE to link keywords to entities (high precision, low recall)
- Method 2: apply mGENRE to map all possible mentions to their entities if exist (higher recall but contains little "Financial Terms")

FinCaKG-Onto Implementations:

merge results from **both** Method 1 and Method 2

3. FinCaKG Construction WITH Ontology
- Phased Results



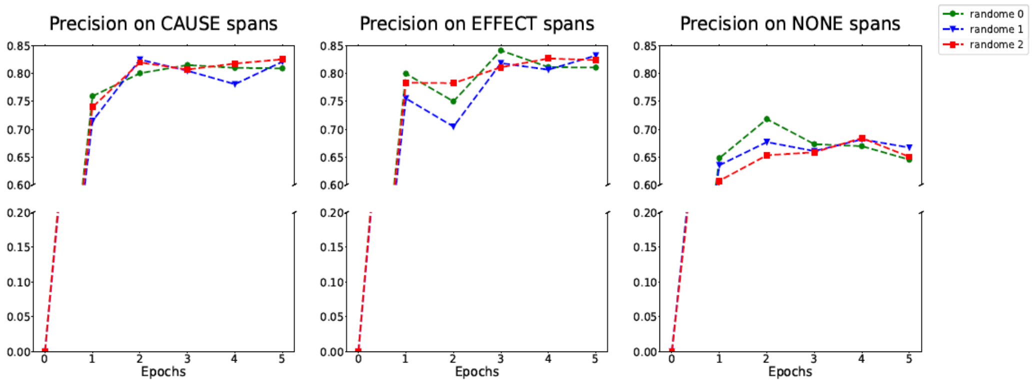


Figure 2: The performance of cause-effect span identification model on test set in Step2.

TABLE III: The performance of causal sentences extraction model in Step1

	Dataset	Precision	Recall	F1-score
Validation	Labeld Data (R1)	96.31%	96.07%	96.19%
Prediction	Financial Reports (R0)	95.93%	95.89%	95.91%

TABLE IV: The performance of financial keywords prediction model in Step3

10-folds	Precision	Recall	F1-score
avg.	98.72%	98.68%	98.70%
min.	98.10%	98.14%	98.12%
max.	99.07%	99.05%	99.06%

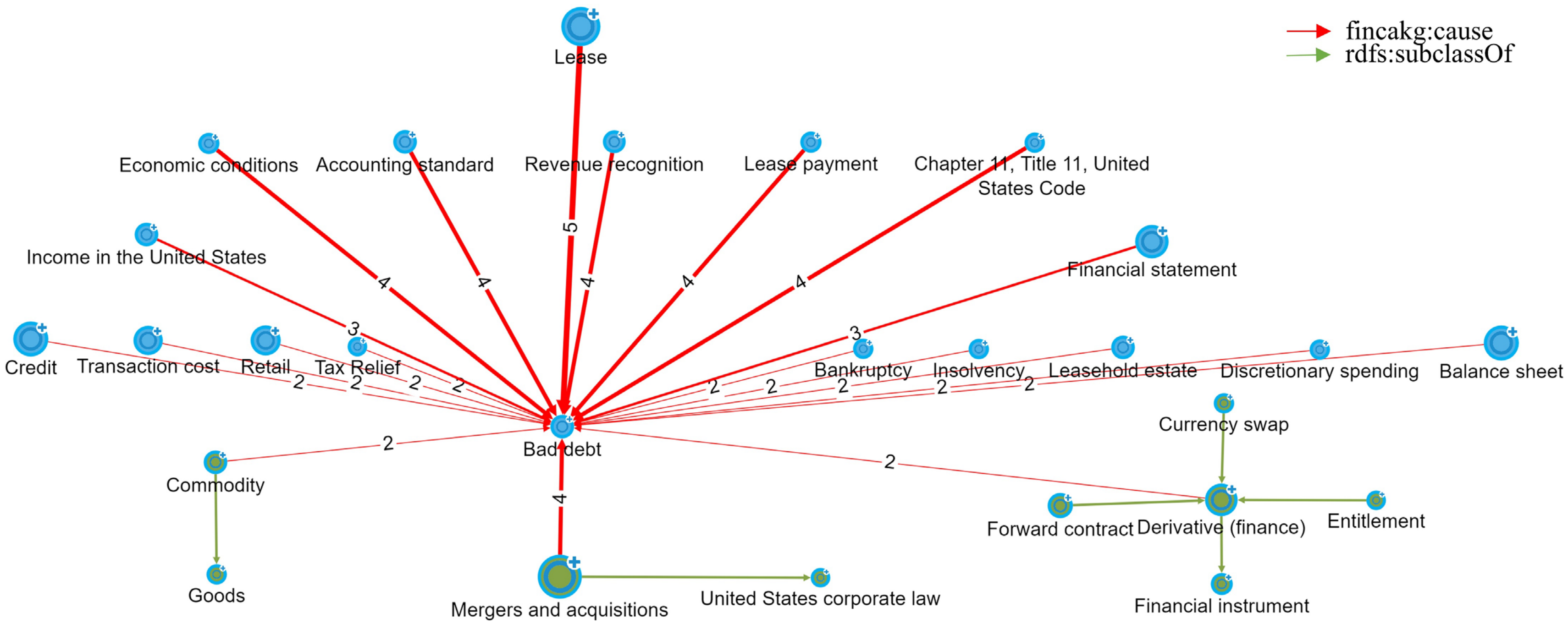
Xu. et al., A Framework to Construct Financial Causality Knowledge Graph from Text, IEEE-ICSC 2024

4. Comparison in Graph Statistics

Table 5: The statistical comparison between distinct causal knowledge graphs.

	FinCaKG	**FinCaKG-Onto**	**Cause@ ConceptNet**
#Docs	5093	5093	-
#Nodes	1317	1554	12832
#Relations	11633	32994	16801
Density	0.39%	0.96%	0.01%
Node Degree	13.55	21.76	2.62

4. Comparison in Inference Exploration



one "fincakg:cause" hop and one "rdfs:subclassOf" hop

5. Conclusion & Limitations

- We illustrated the mono-relation KGs using the FinCaKG framework and incorporate expert ontology into the generation process.
- The statistical comparison between FinCaKG and FinCaKG-Onto indicates the superior performance of FinCaKG-Onto in relation captures
- The incorporation of taxonomic relations through the FIBO ontology opens up exciting possibilities for inferring additional causality in the resulting ordinary FinCaKG.

Limitations:

- We only focus on the finance-related concepts in FinCaKGs, it could **not** cover some **news events**, i.e. Covid eases, energy shock …
- The causal chains might be **less informative** along with **larger hops** in causality

Thanks for your listening and welcome for any questions!