Dynamic Taxonomy and Thematic Filtering for Financial Knowledge Graphs

A. Assadihaghi[1], R. Amini[2], D. Schiemert[3], E. Mackie[4], Z. Xu[5]

Machine Learning Center of Excellence (MLCOE), JP Morgan Chase

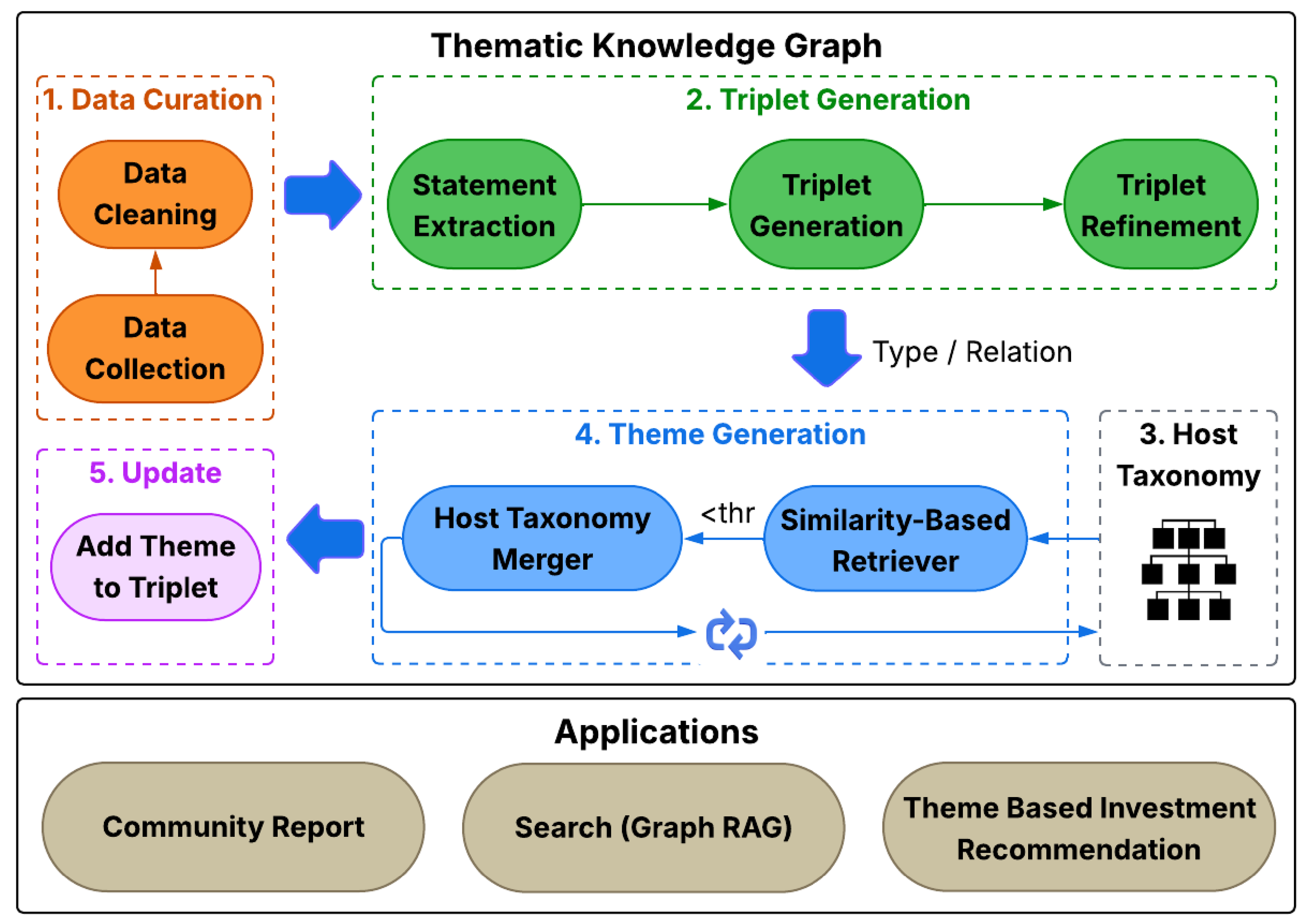
Problem Statement

- Entity-based KGs often suffer from high variability and inconsistency due to evolving financial terminology, therefore, adaptive, high precision structures are helpful
- Our LLM-driven taxonomy and thematic filtering normalize nodes and relations, enabling theme-aware retrieval

Triplet Generation

LLM-based models extract candidate statements from text and convert them into triplets of node, type, and relation

Model	Category	Accurate (%)	Semi-Accurate (%)	Inaccurate (%)
GPT-4o	Head	93	5	2
	Tail	83	13	4
	Relation	82	13	5
	Average	86	10	4
Mistral	Head	72	23	5
	Tail	44	31	25
	Relation	56	26	18
	Average	57	27	16
Flan T5	Head	75	18	7
	Tail	42	16	43
	Relation	39	22	39
	Average	52	18	30



Theme Generation

Taxonomy Update:

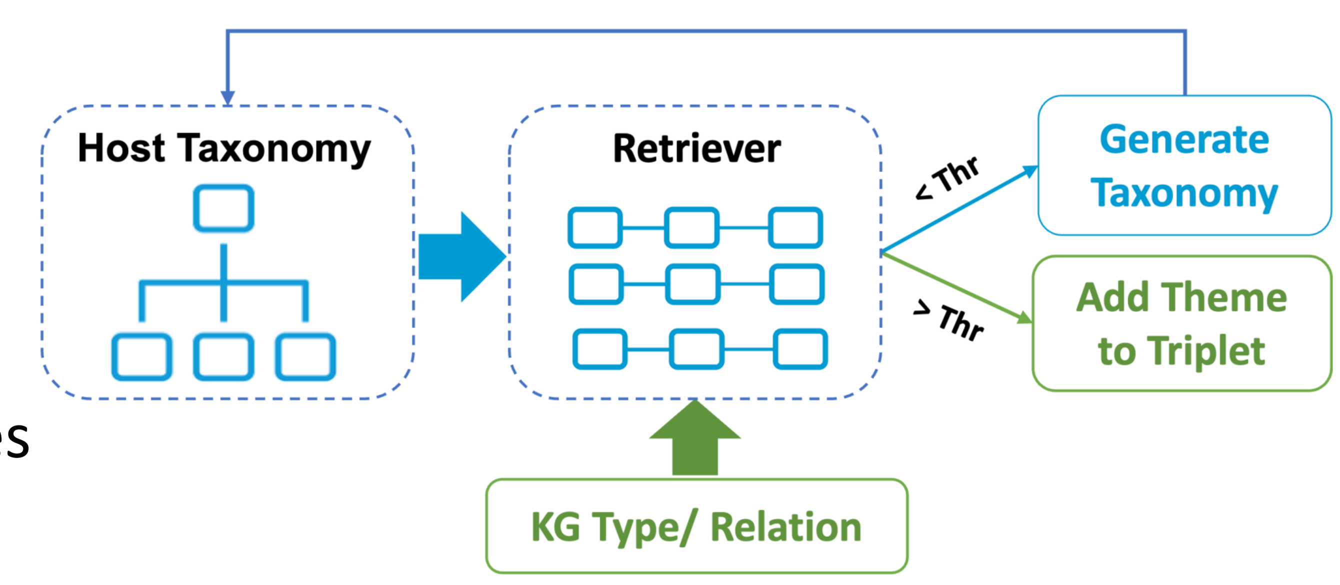
Attach if similarity ≥ threshold (60%); otherwise, review and merge

Taxonomy Generation:

LLM maps surface forms to parent nodes and creates adaptive themes

Taxonomy Merging:

Rooted DAGs with path fuzzy similarity, depth-aware thresholding, and root anchoring



Category	Themes - Grandparent Layer
Relation	Performance, Predict, Policy, Market, Assessment, Causal factor, Strategy, Communication
Type	Resource, Person, Region, Analysis, Strategy, Event, Temporal, Communication, Legal, Process, Investment, Assessment, Economic, Political, Market, Organization, Finance, Policy

Evaluation

Taxonomy Merging:

1,000 nodes are divided into 10 batches

Taxonomy Layers Variability:

1,780 triplets are used for testing

Approach	Misclassification (%)	Missing Nodes (%)	Processing Time (s)
Recursive LLM	8	4	473
Our Approach	4	0	36

Node Property	Grandparent Layer	Parent Layer	Child Layer
Node Type	18	60	2,403
Relation	8	49	944

Use Case: Thematic Information Retrieval

- **Data:** 15,649 finance statements; and 80 questions generated by GPT-4o (each tagged with 1–3 themes)
- **Test:** Retrieval performance compared with and without theme filtering on the same index

Result:

- Theme-based GraphRAG achieves higher F1 scores @1, @5, and @10 compared to standard RAG across most themes
- F1 improvement @1 is statistically significant ($p \approx 0.03$)

Topic	RAG @1	RAG @5	RAG @10	GraphRAG @1	GraphRAG @5	GraphRAG @10
Strategy	0.72 (0.66-0.78)	0.45 (0.36-0.53)	0.25 (0.20-0.29)	0.72 (0.64-0.80)	0.55 (0.47-0.63)	**0.37 (0.32-0.43)**
Predict	0.68 (0.56-0.79)	**0.52 (0.41-0.63)**	0.25 (0.24-0.25)	**0.75 (0.64-0.85)**	0.41 (0.38-0.45)	**0.32 (0.28-0.36)**
Performance	0.37 (0.24-0.50)	0.42 (0.32-0.53)	0.29 (0.21-0.36)	**0.46 (0.35-0.57)**	**0.51 (0.42-0.60)**	**0.37 (0.28-0.46)**
Communication	0.90 (0.87-0.93)	0.75 (0.66-0.85)	0.72 (0.66-0.85)	**0.97 (0.94-1.00)**	**0.88 (0.79-0.96)**	**0.85 (0.74-0.96)**
Causal factor	0.29 (0.20-0.38)	**0.46 (0.41-0.51)**	**0.49 (0.42-0.56)**	**0.35 (0.27-0.42)**	0.42 (0.36-0.48)	0.44 (0.39-0.49)
Assessment	0.42 (0.30-0.55)	0.51 (0.43-0.59)	**0.40 (0.35-0.46)**	**0.45 (0.33-0.56)**	0.51 (0.41-0.60)	0.37 (0.30-0.43)
Policy	0.32 (0.25-0.40)	0.52 (0.43-0.60)	0.47 (0.40-0.55)	**0.35 (0.28-0.41)**	**0.57 (0.50-0.63)**	**0.51 (0.44-0.57)**
Market mvmt	0.49 (0.37-0.61)	0.34 (0.28-0.40)	0.28 (0.21-0.34)	0.49 (0.41-0.57)	0.33 (0.29-0.38)	0.26 (0.20-0.32)
Average	0.55 (0.43-0.67)	0.50 (0.41-0.58)	0.41 (0.34-0.48)	**0.60 (0.48-0.65)**	**0.53 (0.45-0.59)**	**0.45 (0.37-0.50)**

[1] atousa.assadihaghi; [2] reza.amini; [3] daniel.schiemert; [4] ewan.Mackie; [5] zhiwen.xu@jpmchase.com