Parsed 10-K Filing

Financial Lexicon-Aware Mapping (FLAM)
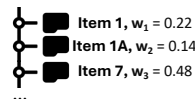
Financial Terms from FinRAD

Semantic Clustering

Weighted Lexicon Mapping

Item 1, w_1 = 0.22
Item 1A, w_2 = 0.14
Item 7, w_3 = 0.48
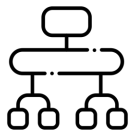...

Lexicon Clustering and Mapping

Assign key terms to disclosure sections using financial lexicon-tuned embeddings. Score semantic similarity to the Items based on relative frequency across filings.

FLAM: Global Knowledge Lexicon Mapping

Lexicons are mapped to disclosure **Items** and assigned weights ($w \in [0, 1]$) based on their relative frequency across filings.

Specifically, financial terms (lexicons) are clustered into latent themes using domain-specific embeddings fine-tuned on FinRAD.
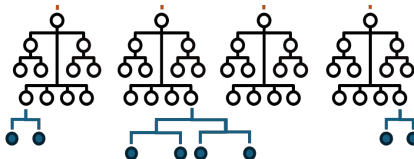
Structure Extraction

SEC Disclosure Tree

One disclosure tree is built for each **Item**, with nodes encoding the semantics of page-level segments.

Hierarchical Expansion

Expanded Hierarchies via Financial Topic Clustering

a. Embed item-level content with financial semantics
b. Cluster into latent disclosure topics
c. Build summary hierarchy over topic clusters
d. Construct question-guided tree from summary tree

Dual-Tree Generation

Dual-Tree Index (Summary + Question)
Summary Tree (Disclosure-Level Indexing);
Question Tree (Query-Driven Retrieval)