



Enhancing Document Retrieval for Technical Knowledge Extraction

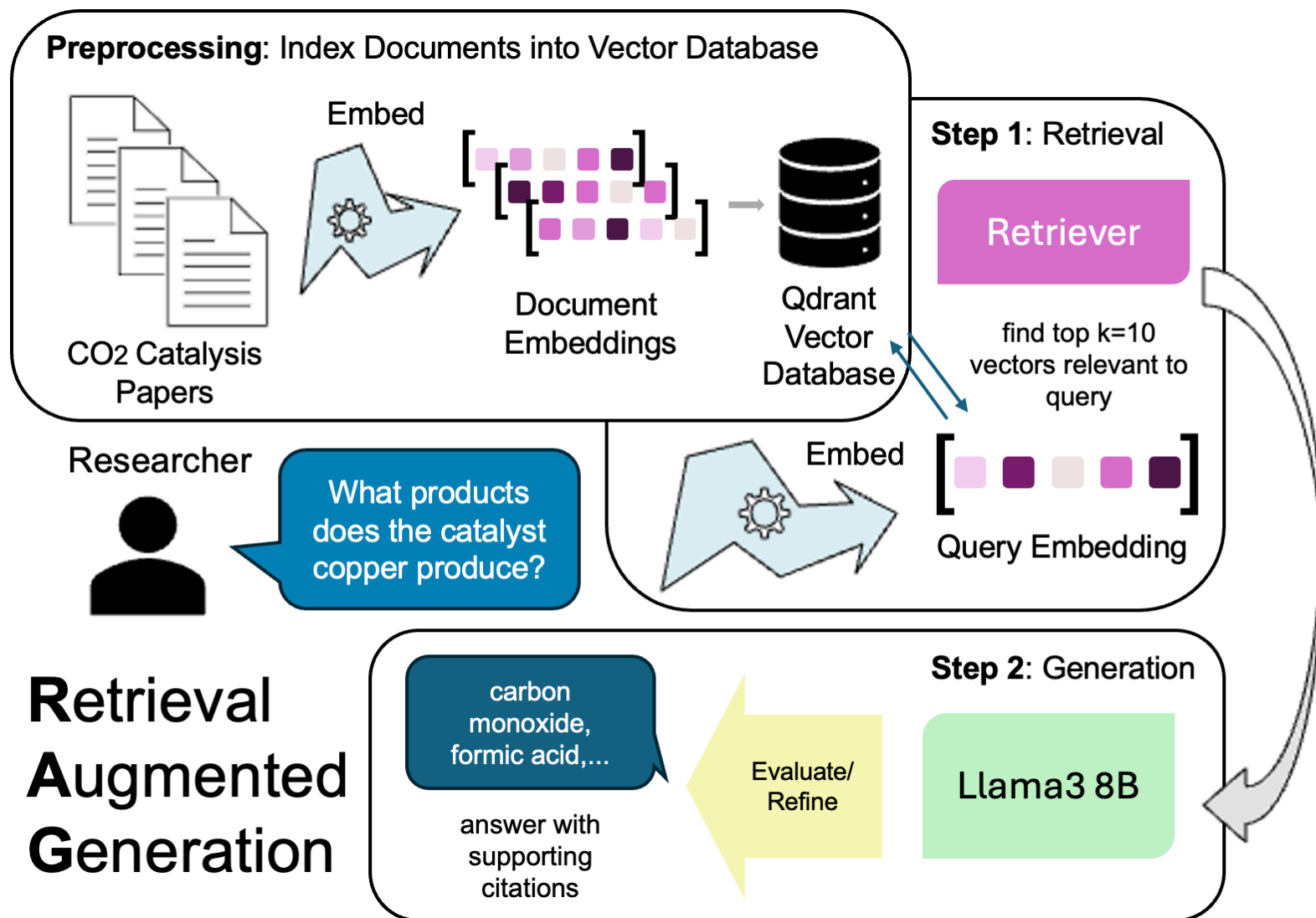
September 11, 2024

Janice Yang

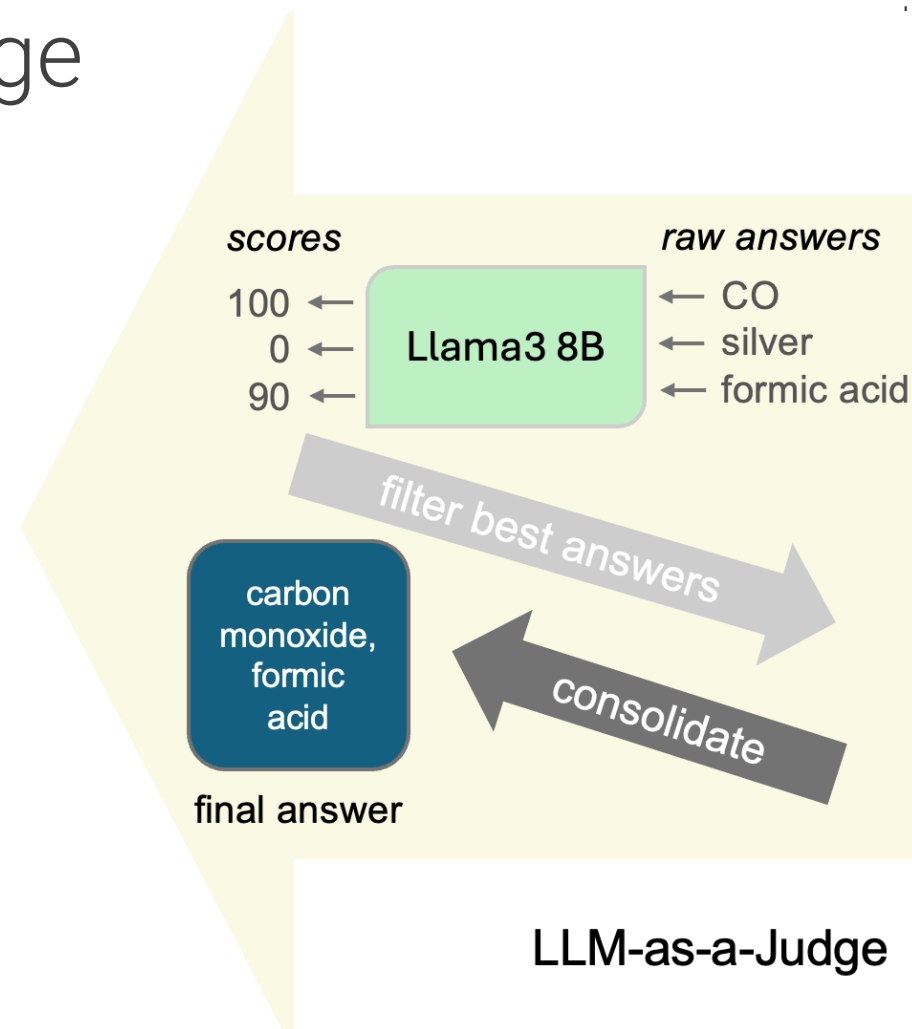
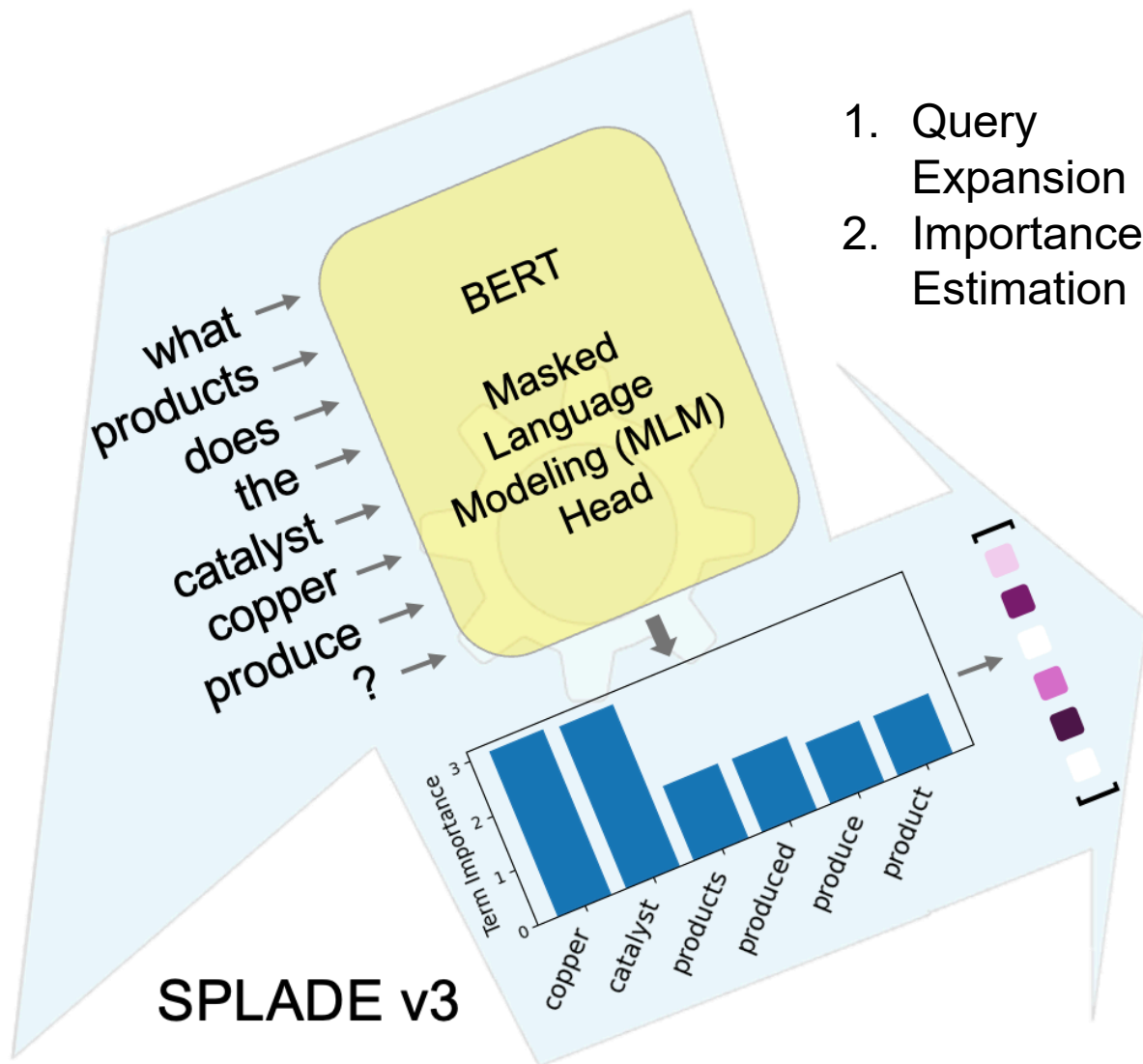


Question-Answering for CO2 Catalysis Research

- **Goal:** enable efficient knowledge extraction from a large document corpus
- **Approach:** Retrieval-Augmented Generation
- **Why this is hard:**
 - LLM hallucination
 - Very specific domain
- **Bottleneck: Step 1, Retrieval**
 - Current: MatBERT dense embeddings + cosine similarity
 - Proposed: SPLADE sparse embeddings



SPLADE-v3 and LLM-as-a-Judge



- evaluates both relevance and groundedness (not correctness)
- often optimistic, but returns 0 for hallucinations

Results and Improvements

References

Example Results

Query

What products does the catalyst copper produce?

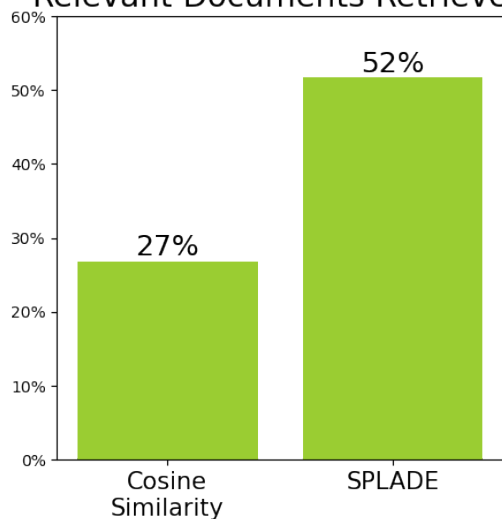
Answers

methane [1], CO [1, 6],
carbazoles [2], iodide [2],
elemental iodine [2], formic
acid [3, 5], alcohols [3, 4, 5],
hydrocarbons [4, 5], CH₂ [6],
CH₄ [7], C₂H₄ [7], H₂O [8],
CH₃OH [8], CO₂ [9]

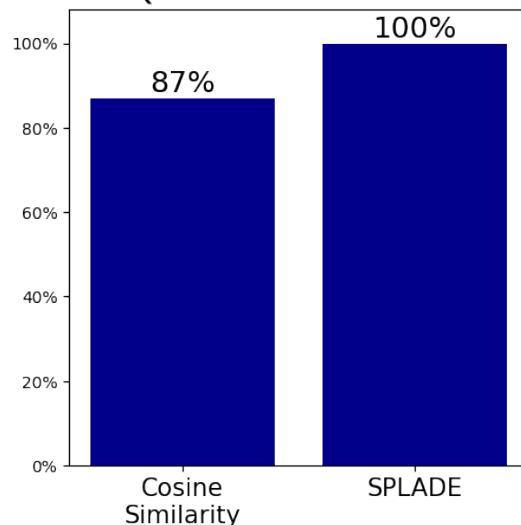
- [1] Although heterogeneous catalysts such as copper can produce highly reduced products, including methane (8), ethanol (12), and ethylene (10), the complexity of the surfaces on which these reactions take place and number of products makes detailed mechanistic studies difficult.
- [2] The excited copper catalyst undergoes SET with molecular iodine to produce iodide and elemental iodine.
- [3] Copper-based catalysts have been proved to show excellent performance in electrochemical CO₂ reduction reaction (CO₂RR), which can reduce CO₂ to a variety of value-added liquid products like formic acid and alcohols [22, 25].
- [4] It has always been a barrier to study the selective production of a single product, or a mixture mainly composed of the desired product, in a copper-based catalyst system.
- ...

Comparison to previous method

Relevant Documents Retrieved



Questions Answered



Determining utility for SME through QA correctness comparison with and w/o RAG tools

CO₂ Quiz Clean

1. What products does the catalyst copper produce?
2. What products does the catalyst gold produce?
3. What other catalysts are like gold, silver, and copper?

CO₂ Quiz (with AI-assistant)

1. What products does the catalyst copper produce? *

AI-assistant:
methane [1], ethanol [1], ethylene [1], CO [1, 6, 8], formate [1], carbazoles [2], radical cation [2], iodide [2], elemental iodine [2], formic acid [3, 5], alcohols [3, 4, 5], hydrocarbons [4, 5], CH₂ [6], CH₄ [7], C₂H₄ [7], H₂O [8], CH₃OH [8], H₂ [8], CO₂ [9]

References:
[1] Although heterogeneous catalysts such as copper can produce highly reduced products, including methane (8 e.Ai), ethanol (12 e.Ai), and ethylene (10 e.Ai).4, 5 the complexity of the surfaces on which these reactions take place and number of products makes detailed mechanistic studies difficult.
[2] The excited copper catalyst undergoes SET with molecular iodine to produce iodide and elemental iodine.
[3] Copper-based catalysts have been proved to show excellent performance in electrochemical CO₂ reduction reaction (CO₂RR), which can reduce CO₂ to a variety of value-added liquid products like formic acid and alcohols.
[4] It has always been a barrier to study the selective production of a single product, or a mixture mainly composed of the desired product, in a copper-based catalyst system.
[5] In hetero- gogeneous catalysis, copper and copper oxide catalysts can selectively produce hydro- carbons from CO₂, but the efficiency is relatively low due to the high overpotential.
[6] To sum up, besides sintering, copper catalysts were subjected to carbonaceous products fouling which block the active sites of the catalyst, thus reducing its catalytic activity.
[7] Copper is an adequately researched material for use as a catalyst in the multi-electron reduction process leading up to the production of methane (CH₄), ethane (C₂H₆), and methanol (CH₃OH).
[8] For instance, copper-based catalysts in the dura- tion of methanol conversion at small pressure produce also CO, which seriously poisons the active components on the surface of the catalyst [12, 14].