



Ismail Berbache



Alessandro Morosini

Turning Emails into Knowledge

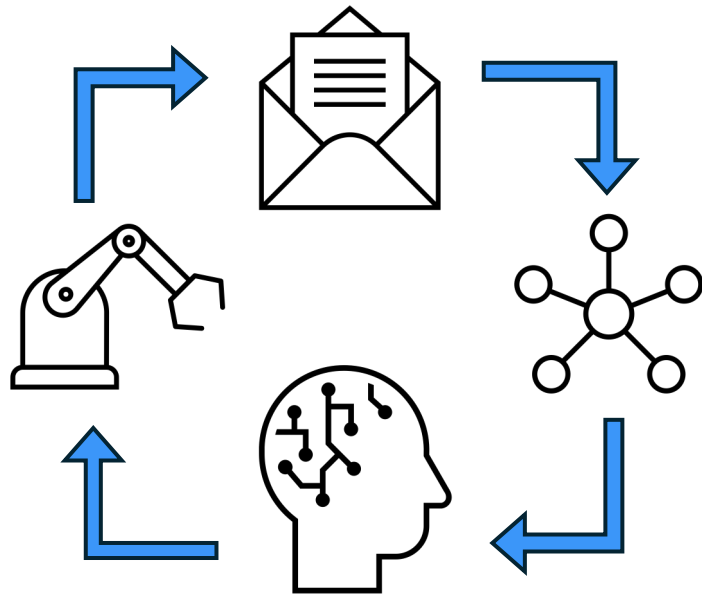
Company Advisors: Ian Campbell, Peter Greskoff, Grace An Faculty Advisor: Ilya Jackson

Project Overview

Problem Statement

Lineage spends **over 1M** hours every year dealing with customer emails, covering tasks like rescheduling or order changes. Yet key **information remain locked in free text**.

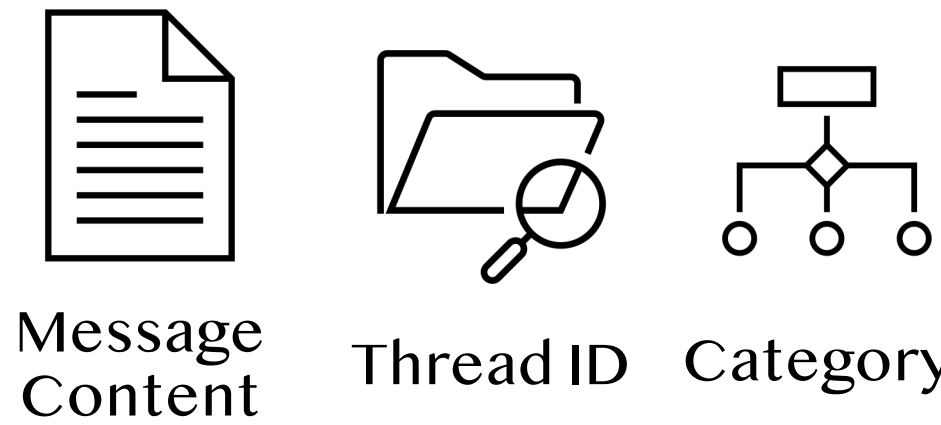
To unlock insights from these emails and automate repetitive customer service tasks, we extract **structured knowledge** from the text.



Data

- Millions of customer service emails
- Threads reconstructed to capture full exchanges
- 7 distinct categories
- Presents challenges (artifacts, shuffling...)

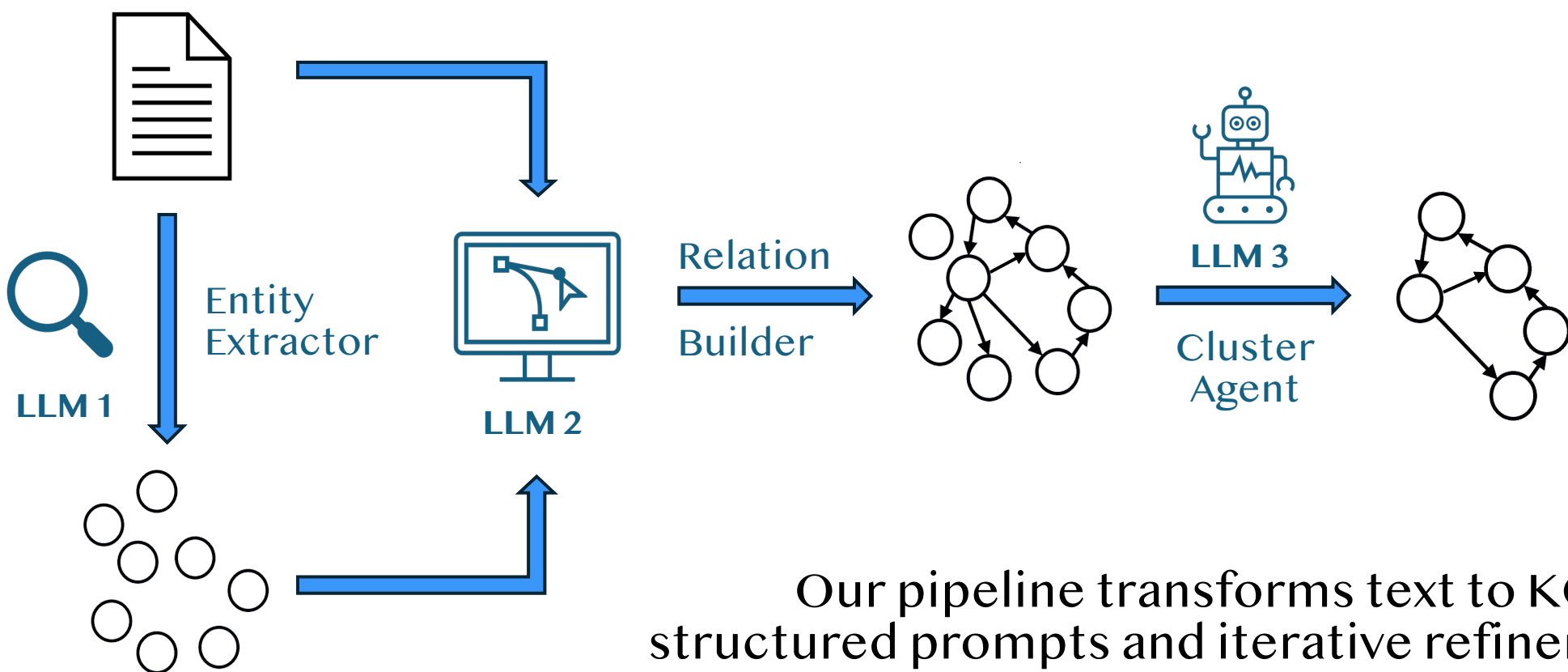
3 key fields



Turning Conversations into Graphs

Concrete Knowledge Graph (KG) Extraction

- Goal:** Represent each email thread as a **graph** of entities and relationships
- Nodes:** Specific entities mentioned (e.g. PO #123, next Monday)
- Edges:** Semantic links capturing actions (e.g. asks to reschedule, confirm)

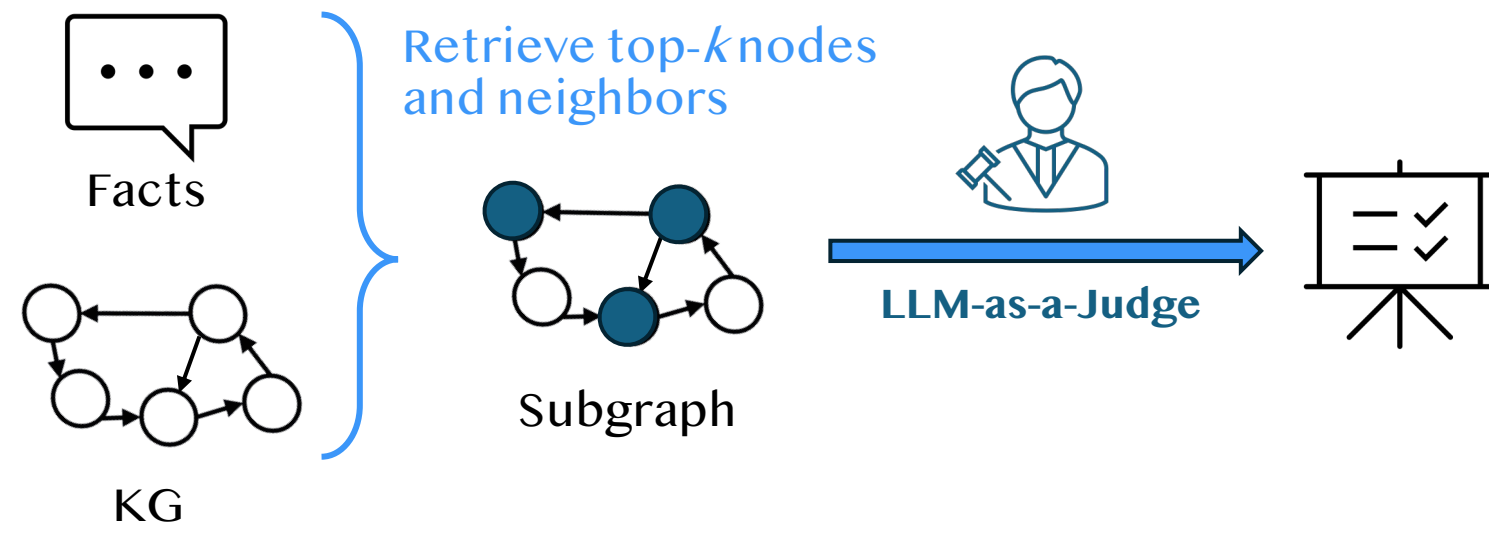


Our pipeline transforms text to KGs via structured prompts and iterative refinement.

Evaluation

- Evaluating KGs is complex due to ambiguous ground truth
- We built **Lin-MINE** to evaluate how well graph extraction pipelines capture information from Lineage emails

Idea: Check if key facts from an email can be found within its KG



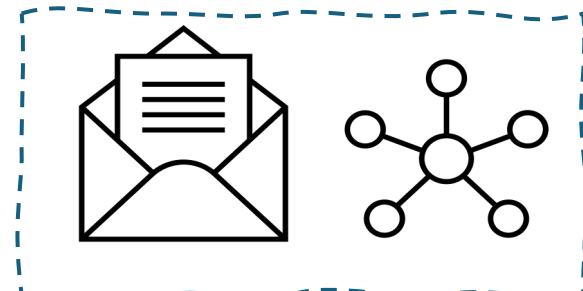
Output: For each email, % of facts in KG

Abstracting Graphs into Patterns

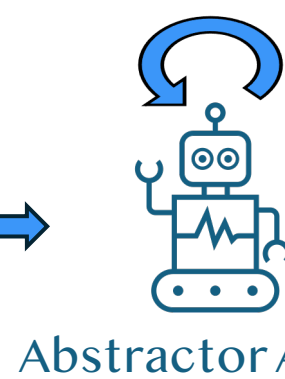
Abstraction

- Goal:** Map concrete mentions to a **high-level** representation
- Method:** Chain of LLMs performs context-aware abstraction and normalization
- Output:** Abstract KG with normalized nodes for each thread

Email and Concrete Entities

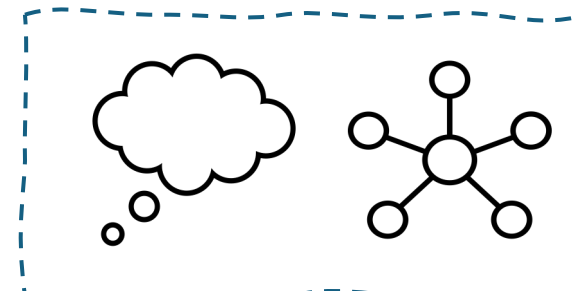


E.g. PO #123

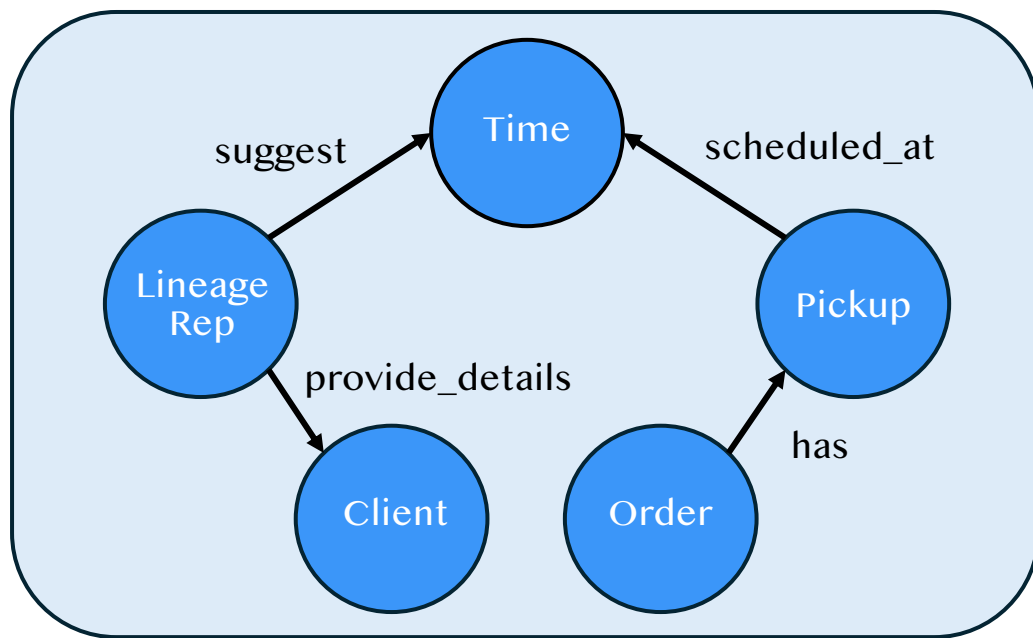
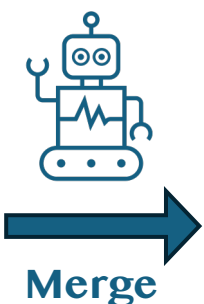
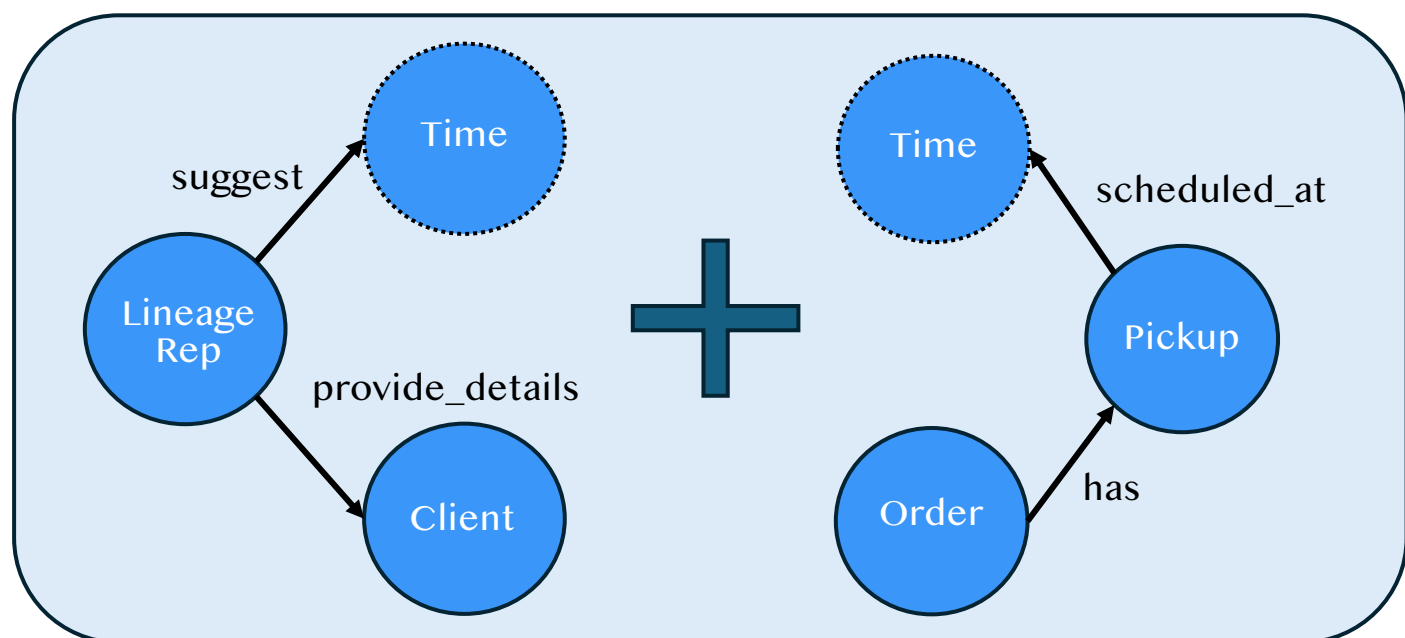


E.g. Order, Reference Number, PO ID

Normalized Abstract Entities



E.g. Purchase Order



Merging

- Combine abstract graphs from multiple threads
- Identify recurring patterns across conversations
- Produce category-level schemas and insights
- Enable automation and downstream tasks

Deliverables & Business Impact

Deliverables

70%+

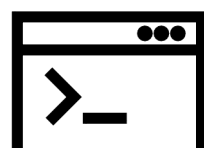
Key facts captured in structured way



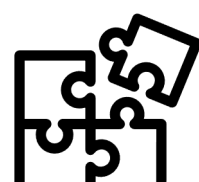
Custom eval built for Lineage emails



3 layers of abstraction

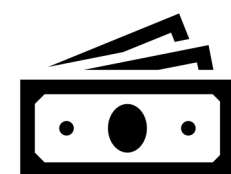


User-focused, terminal-ready code



Plug-and-play LLM components

Estimated Annual Impact



\$12 - 24M

Payroll savings