



Gieni Lookup Bot

Orderfox Challenge Datathon 2025

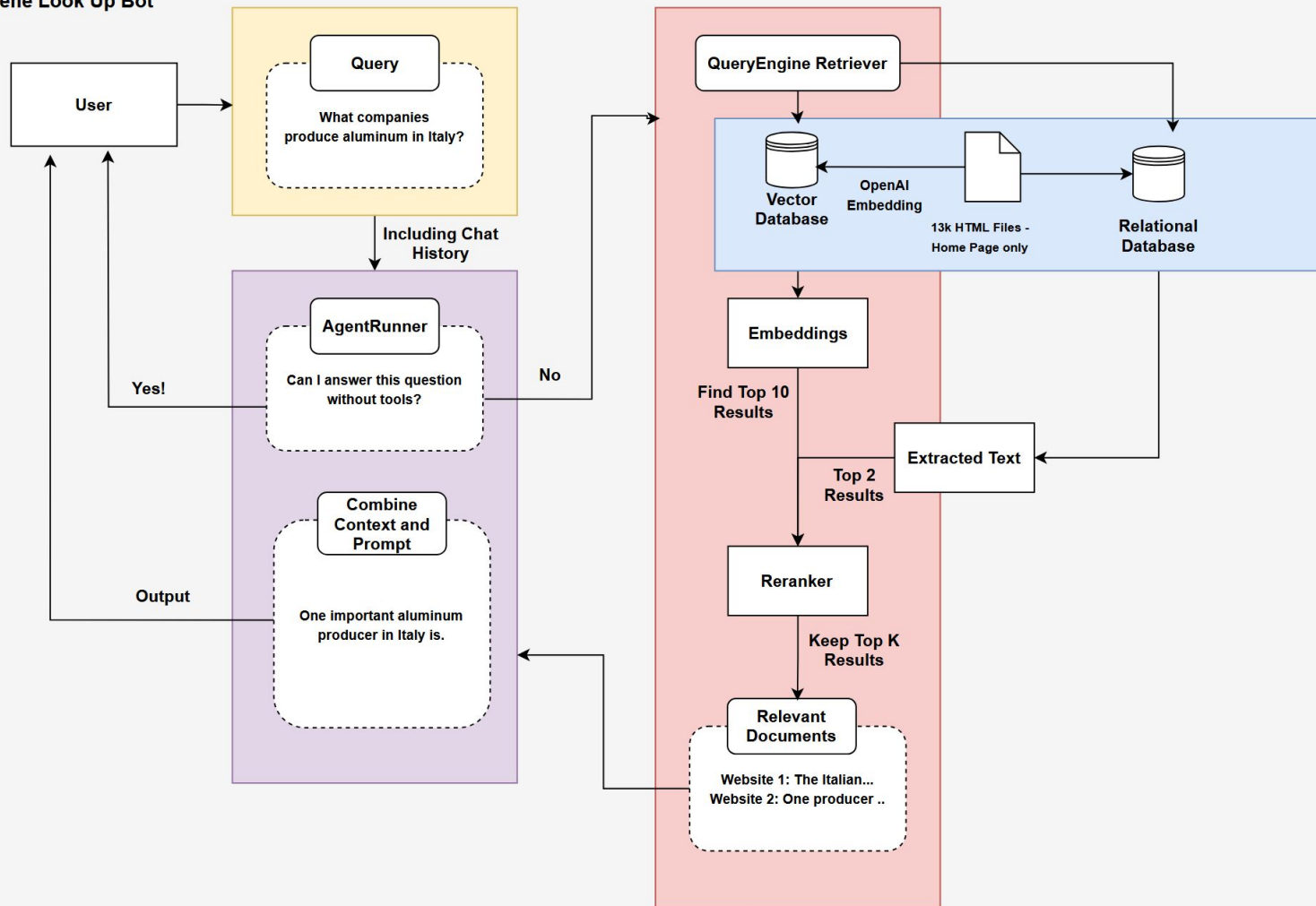
Hanno Hiss, Philippe Zwick, Laurenz Zehetner, Maximilian Scheidl



Data Preprocessing

- 13.000+ Websites each containing 58 subpages on average
- Identified Problems
 - CSS files
 - Duplicated pages
 - Images and Video embedded as text
 - Redundant text (Imprint is on every page)
- Solution
 - Removing CSS files reduced around 10% of redundant text
 - Cutting all pages which have more than 50.000 Characters to avoid images in

Giene Look Up Bot



Workflow



1. Add API Key



Model Parameters

Configure the settings for the chatbot below.

Top k results

5

-

+

Max Tokens

500

-

+

Enter API Key:

.....



Submit API Key & start with this configuration



Gieni Lookup Bot

Please enter your API key to start the chatbot.

2. Greet Politely



Gieni Lookup Bot



Hi! How are you?



Hello! I'm just a program, so I don't have feelings, but I'm here and ready to help you. How can I assist you today?

3. Ask your Questions



I am concerned about geopolitical risk in my supply chain. In which region should i look for alternative specialty chemical suppliers?



To mitigate geopolitical risk in your supply chain for specialty chemicals, consider looking for alternative suppliers in the following regions:

1. **North America:** The United States and Canada have robust chemical manufacturing sectors and a diverse supply chain network.
2. **Western Europe:** Countries like Germany, the Netherlands, and France are known for their advanced chemical industries and strong regulatory frameworks.
3. **Southeast Asia:** Nations such as Singapore and Malaysia offer growing chemical markets with favorable trade agreements and logistics infrastructure.
4. **Latin America:** Brazil and Mexico have emerging chemical sectors and can serve as alternative sources.
5. **Australia and New Zealand:** These countries provide stable environments and are increasingly involved in specialty chemical production.

Exploring suppliers in these regions can help diversify your supply chain and reduce reliance on any single geopolitical area. If you need more specific information or assistance, feel free to ask!



4. Don't forget to say Thank You (once)



Great! Thank you



You're welcome! If you have any more questions or need further assistance, just let me know.
Have a wonderful day!



Future Ideas

- Active Fine-Tuning
 - To better address the relevance problem, one can use test time fine tuning to optimize information gain about the query rather than naive similarity search.
- Utilize the whole data
 - Either embed all with Open AI and wait for some days or use smaller models for faster ingestion at the cost of retrieval accuracy
- Metadata Extraction
 - Using LLMs to scrape structured metadata from all pages to further improve retrieval
- More Data Cleaning
 - Removing redundant text or summarize pages to their main contents to reduce noise