Kaggle Comp.

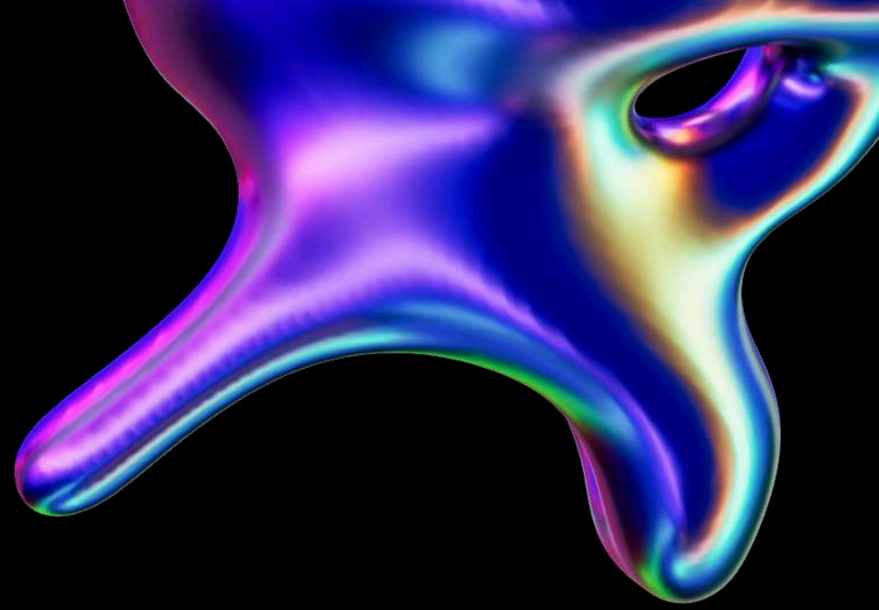
Hackapizza

Community Edition

Bytebusters
Stefano Iannicelli
Ettore Caputo

Hackapizza
6 marzo, 2025

Panoramica della soluzione



Main idea

Data preparation

Architecture

Keyword Extractor

Tech Expert

Distance Expert

Menu Header Expert

Menu Corpus Expert

Criticità e possibili miglioramenti

Main idea

Estrazione delle keyword,
riformulazione con LLM specializzati,
ricerca mirata.

Certainty

Desideravamo garantire che i chunk di testo impiegati nel processo di RAG fossero esclusivamente quelli realmente pertinenti alla domanda.

Token efficiency

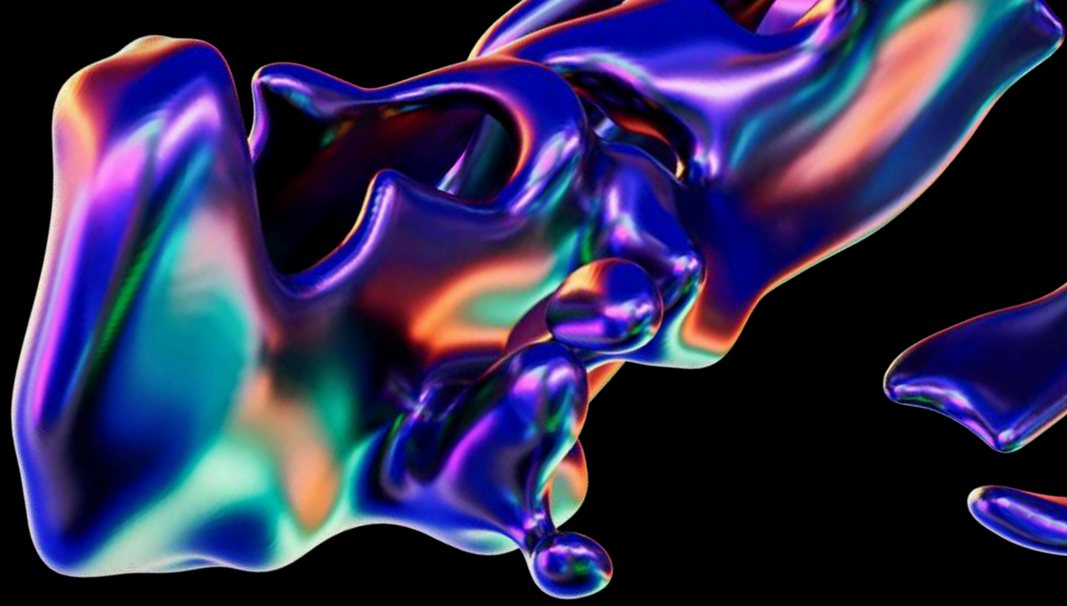
Sfruttando la certezza sui chunck, abbiamo rimosso l'utilizzo del LLM durante l'estrazione del nome dei piatti dai chunk. Risparmiando sui token di contesto.

Task distribution

Il processo di generazione della risposta è articolato in più fasi, ciascuna gestita da diversi LLM "specializzati".

Data preparation

Trasformazione e formattazione



1

Split dei file menu

Abbiamo suddiviso ogni file menù in:

- header (la porzione che precede la lista dei piatti)
- un chunck per ogni piatto contenuto nel menù.

2

Etichettatura dei piatti

All'interno dei file, i nomi dei piatti sono stati racchiusi tra i tag <dish> </dish>. Questo ne ha reso più semplice l'estrazione dal contesto.

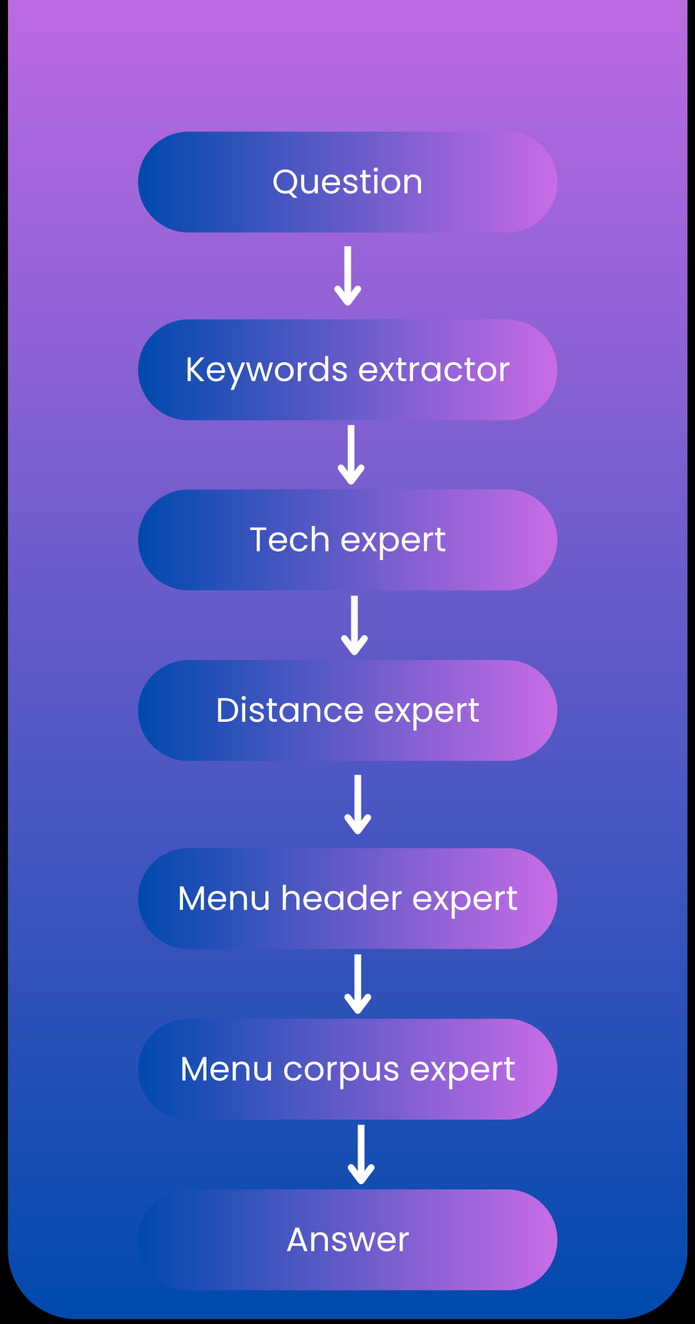
3

Tabelle e altro

Le tabelle sono state ricostruite poiché la conversione da PDF a TXT le aveva corrotte. I numeri romani sono stati trasformati mediante l'uso di una regex.

Architecture

- Dalla domada vengono estratte tutte le keywords.
- Le keywords vengono utilizzate dai vari esperti.
- Il Tech expert ottiene informazioni sulle tecniche presenti nel Codice di Galattico.
- Il Distance expert ottiene informazioni sulle distanze tra pianeti.
- Il Menu header expert ottiene informazioni dai soli header dei menu.
- Il Menu corpus expert ottiene informazioni dai piatti contenuti nei menu.
- La risposta viene estratta dal contesto fornito dal Menu corpus expert.



Keyword extractor

Prompt ad un LLM

Viene chiesto ad un LLM di estrarre le keyword dalla domanda, cotegorizzandole come segue:

```
{
    "licenze chef": [ ... ],
    "licenze tech": [ ... ],
    "pianeti": [ ... ],
    "ingredienti": [ ... ],
    "tecniche galattiche": [ ... ],
    "tecniche comuni": [ ... ],
    "chef": [ ... ],
    "ristoranti": [ ... ]
}
```

Tech expert

Attivazione

Questo esperto si attiva solo quando nelle tra keyword estratte ci sono le “licenze tech” o “tecniche comuni”.

Estrazione delle tecniche

Quando utilizzato, il tech expert cerca le tecniche presenti nel Codice Galattico, che soddisfano la query sulla base delle keywords di attivazione.

Per fare ciò diamo in input il Codice Galattico come contesto al LLM.

Quindi, viene riformulata la query sulla base delle nuove informazioni ottenute, aggiungendo le tecniche estratte dal Codice Galattico.

Esempio

Quali piatti utilizzano tecniche di congelamento non contengono Teste di Idra?

Quali piatti utilizzano il “Congelamento Quantico” e il “Congelamento Supersonico” ma non contengono Teste di Idra?

Distance expert

Attivazione

Questo esperto si attiva solo se nella domanda è presente “anni luce”.

Estrazione dei pianeti

Quando utilizzato, il distance expert cerca i pianeti che soddisfano la query.

Per fare ciò diamo in input il file delle distanze nel contesto del LLM.

Quindi, viene riformulata la query sulla base delle nuove informazioni ottenute, aggiungendo i pianeti che rispettano la distanza richiesta.

Esempio

Quali piatti preparati in ristoranti distanti 300 anni luce da Namecc contegono Teste di Idra?

Quali piatti preparati su Asgard e Pandora contegono teste di idra?

Header menu expert

Attivazione

Questo esperto si attiva solo se sono presenti keyword di:

- "ristoranti"
- "chef"
- "pianeti"
- "licenze chef"

Poichè nell'header sono presenti al più queste informazioni.

Estrazione degli id dei documenti

Quando utilizzato, questo nodo trova gli id dei menu che rispettano la query.

Per fare ciò chiediamo ad un LLM di riformulare la query in una query booleana con solo le keyword di attivazione nodo.

Una volta che il modello ha formulato la query booleana eseguiamo una ricerca su tutti gli header dei menu e ritorniamo solo gli id che rispettano la query booleana.

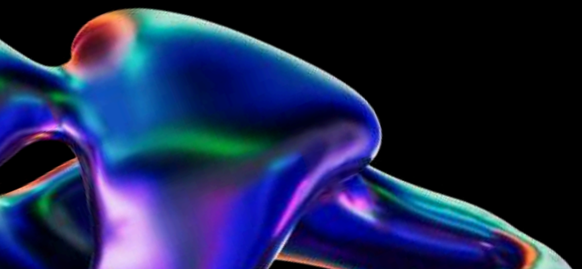
Esempio

Quali piatti preparati in ristoranti sul pianeta Pandora dallo chef Cannavacciuolo contengono Lumache Lucenti?

Keyword da usare per formare la query booleana: ["Pandora", "Cannavacciuolo"]

Output LLM: ("Pandora" AND "Cannavacciuolo")

Output node: [13, 16, 7] 13,16,7 sono id di ristoranti tale che il loro header rispetta la query booleana



Corpus menu expert

Attivazione

Questo esperto si attiva solo se sono presenti keyword di:

- “tecniche galattiche”
- “ingredienti”

Arrivati a questo nodo, se i precenti hanno funzionato esiste almeno una keyword di “tecniche galattiche” o “ingredienti”

Estrazione degli id dei documenti

Quando utilizzato, questo nodo trova i chunk di testo che contengono i nomi dei piatti che rispondono alla query.

Per fare ciò chiediamo ad un LLM di riformulare la query in una query booleana con solo le keyword di attivazione nodo.

Una volta che il modello ha formulato la query booleana eseguiamo una ricerca su tutti i corpus dei menu e ritorniamo solo i chunk che rispettano la query booleana.

Esempio

Quali piatti preparati tramite la tecnica di Congelamento Quantico utilizzano Spore Radioattive ma non le Alghe Fritte?

Keyword da usare per formare la query booleana: [”Congelamento Quantico”, “Spore Radioattive”, “Alghe Fritte”]

Output LLM: (“Congelamento Quantico” AND “Spore Radioattive” AND NOT “Alghe Fritte”)

Output node: La lista dei chunk di testo che contengono i piatti richesti nella domanda iniziale.

Extract answer

Estrazione piatti

Prima di effettuare l’estrazione, viene fatta l’intersezione tra gli id dei ristoranti che hanno soddisfatto le query booleane degli esperti “header” e “corpus”.

Vengono quindi estratti tutti i nomi dei piatti presenti nei chunk ottenuti dopo l’intersezione.

Grazie alla fase di etichettatura iniziale, i piatti nel testo sono facilmente estraibili poiché tra i tag <dish>< \dish>. Quindi

RAG!!! not so much

Poichè sono state utilizzate query booleane abbiamo la certezza che i chunk ritornati dal menu corpus expert siano tutti e solo i chunk che contengono nomi di piatti che rispondono alla domanda iniziale. Per questo motivo non c’è bisogno di usare un LLM per individuare le risposte.

Risultati ottenuti

76.465

Only Menu Expert

63.5

score

Adding Distance Expert

66.7

score

Adding Tech Expert

76.465

score

Analisi
critica della soluzione proposta

Modello booleano rigido:

Le query booleane che costruiamo durante il processo sono notoriamente rigide. Errori nei nomi delle keyword non sono ammessi.

Implementazione del tech expert

Nel tech expert si da' in input all'LLM tutto il Codice Galattico. Un notevole risparmio di token si otterrebbe costruendo un sistema che restituisca solo i chunk che contengono effettivamente la risposta.

P.S. purtroppo non abbiamo avuto molto tempo a disposizione, entrambi stiamo scrivendo la tesi magistrale e la scadenza è incombente

Grazie

stefano.iannicelli362@gmail.com

ettore.caputo27@gmail.com