



Enhancing Location-Based Food and Beverage Responses with Open Source LLMs

Project Report

50.045 Information Retrieval

Team Members:

Mohammed Fauzaan (1005404)
Rohit Raghuram Murarishetti (1005398)
Ishan Monnappa Kodira (1005284)
Lee Wai Shun (1005115)

1 Executive Summary

The report outlines a comprehensive information retrieval system designed for restaurant recommendations. The architecture involves a main pipeline with components like Query Rewriter, Annotate Needs, Search Term Generator, and more. Elasticsearch is utilized, driven by the BM25 algorithm, to retrieve relevant restaurant information. Additionally, a Google Evaluation Pipeline is employed for result comparison. The project integrates the vicuna-7b-v1.5 LLM model for tasks such as generating clarifying questions, search queries, and restaurant recommendations. The LLM instances are tailored for different stages of the pipeline.

The evaluation process involves assessing the system's performance against ground truth data obtained through Google's BARD and manual fact-checking. Various metrics, including Jaccard Similarity and Mean Average Precision, are employed to gauge the system's precision. The analysis suggests room for improvement, particularly in data source expansion and algorithmic enhancements.

The precision analysis based on a proxy ground truth using Google Maps data reveals challenges in aligning with Google's retrieval methodologies. The report concludes with suggested improvements, emphasizing the need for data source expansion and exploring advanced retrieval algorithms.

Contents

1	Executive Summary	1
2	Problem Statement	4
2.1	Problem Definition	4
2.2	Code Repository	4
3	Significance and Motivation	4
3.1	Expected Outcomes of the Project	4
4	Query Analysis and Structure	5
4.1	Geographical Context	5
4.2	Restaurant Properties	5
4.3	Query Examples	5
4.4	Idea Generation and Encoding Queries	5
4.5	Complex Queries	6
4.6	Purpose and Application	6
5	Challenges	6
5.1	Diversity of Cuisines	6
5.2	Location-Based Relevance	6
5.3	Contextual and Personalized Responses	6
6	Literature Review	6
6.1	Query Rewriter	6
6.2	Retriever	7
6.3	Reranker	7
6.4	Reader	7
6.5	Integration	7
7	Dataset	8
7.1	Exploration 1: Web Extracted Data	8
7.1.1	Limitations of Web Extracted Data	8
7.2	Exploration 2: Yelp API	9
7.2.1	Limitations of Yelp API	10
7.3	Chosen: Google Places API	11
8	Overall Architecture	12
8.1	Main Pipeline	12
8.2	Google Evaluation Pipeline	14
9	Retrieval	14
10	Generation	15
10.1	LLM Usage:	15
10.1.1	Query Rewriter	16
10.1.2	Annotate Needs	16
10.1.3	Search Term Generator	16

10.1.4	Reader	16
11	Verification and Citation Generation	17
11.1	Creating Annotations for LLM Output Based on Context	17
11.1.1	BERTScore for Semantic Similarity Assessment	17
11.1.2	Matching Tokens	18
11.1.3	LLM-based Judgement	18
11.2	Selection	21
12	Evaluation	21
12.1	Baseline LLM Generation	23
12.2	Analysis against SOTA-Generated Answers	23
12.3	Precision Analysis based on Proxy Ground Truth	25
12.4	Evaluation Repository	26
12.5	Suggested Improvements	26
A	References for Generated Outputs	28
A.1	Entry 1	28
A.2	Entry 2	30
A.3	Entry 3	34
A.4	Irrelevant Context	37
B	Proxy Precision	39

2 Problem Statement

2.1 Problem Definition

Our project aims at enhancing responses for location-based Food and Beverage (F&B) queries using Open Source LLMs, specifically Vicuna 7B. Specifically, we aim to augment the generated location-based responses for diverse F&B queries across various cuisines, ensuring the relevance, accuracy, and informativeness of the output.

2.2 Code Repository

The code repository for this project can be accessed by clicking through this link:

www.github.com/waishun78/hungry-rag

3 Significance and Motivation

This project addresses the problem of outdated and subpar responses generated by Open Source LLMs when catering to location-specific F&B queries. As individuals increasingly rely on digital platforms for dining choices, there is a growing demand for precise and contextually relevant information regarding restaurants and cuisines in specific areas. Enhancing these responses with traditional information retrieval systems can create a better experience for consumers.

3.1 Expected Outcomes of the Project

The goal of this project is to develop a Retrieval Augmented Generation (RAG) model that excels in delivering personalized, location-specific food and beverage recommendations. The anticipated outcomes include:

- **Enhanced Accuracy and Relevance:** The system should consistently provide accurate suggestions that align with the diverse range of cuisines and specific location-based requirements, thereby catering to a global audience with varied culinary interests.
- **Contextual Comprehension and Personalization:** A key objective is to ensure that the model can interpret and respond to nuanced aspects of queries. This involves an understanding of personal preferences, dietary needs, and specific dining environments, thereby enhancing the user experience.
- **Utility for Consumers and Businesses:** The model is expected to serve as a valuable tool not only for individuals seeking tailored dining suggestions but also for local restaurants and food businesses aiming to connect with their target audience effectively.

In essence, this project aims to bridge the gap between user expectations and the capabilities of existing language models, elevating the overall quality of location-based F&B recommendations.

4 Query Analysis and Structure

We analysed the query structure for restaurant recommendations to accommodate a wide range of user preferences and situations. It addresses two main categories: geographical context and restaurant properties.

4.1 Geographical Context

The geographical context includes parameters such as proximity to landmarks, specific neighborhoods, and scenic views. Examples of these queries are:

- Near [landmark]
- In [neighborhood]
- Within walking distance of [attraction]
- With a view of [city skyline/waterfront/etc.]

4.2 Restaurant Properties

Restaurant properties focus on specifics like group size accommodation, opening hours, family-friendliness, atmosphere, pricing, and dietary accommodations. Examples include:

- Group size: For [number] people
- Opening hours: Open late, on weekends, for breakfast, etc.
- Family friendly: Kids' menu, high chairs, changing tables
- Setting/Atmosphere: Casual, fine-dining, romantic
- Price: Budget-friendly, mid-range, fine-dining
- Dietary Needs: Vegan, vegetarian, gluten-free, nut-free options
- Dietary Restrictions: Halal, Kosher Friendly

4.3 Query Examples

We designed queries to combine these elements for comprehensive search results. Examples include:

`Where to get Italian food in Bugis that is family-friendly and has a casual setting?`

`Where to get Japanese food in Marina Bay Sands that is fine-dining and has a romantic atmosphere?`

4.4 Idea Generation and Encoding Queries

This structure also supports idea generation and encoding queries for specific events or activities. For instance:

`I'm looking for a casual restaurant for a date night. What are some suggestions?`

4.5 Complex Queries

The system also handles complex queries involving multiple targets or varied requirements, such as:

```
I want to buy bubble tea and then have Korean fried chicken. Can you  
give me some suggestions?
```

4.6 Purpose and Application

The purpose of this structured query approach is to provide users with tailored restaurant recommendations that align with their specific needs and preferences, enhancing the overall experience and satisfaction.

5 Challenges

5.1 Diversity of Cuisines

One of the primary challenges lies in enabling the Retrieval Augmented Generation (RAG) system to accurately respond to inquiries encompassing a broad spectrum of global cuisines and tastes. Ensuring the relevance and cultural authenticity of the recommendations remains a key challenge.

5.2 Location-Based Relevance

Another significant challenge is the integration of location-aware responses. The system should not only recognize the geographical context of the query but also align its recommendations with the local culinary landscape. This involves understanding and incorporating details about local dining hotspots, proximity to landmarks or attractions, and regional dining preferences.

5.3 Contextual and Personalized Responses

The intricacy of queries demands a sophisticated understanding of various factors, such as individual dietary restrictions, preferred dining settings, and specific meal times. The system must interpret these multifaceted requirements and preferences to provide personalized and contextually relevant dining suggestions.

6 Literature Review

In our research endeavor, we thoroughly investigated various implementation strategies and pipeline architectures. Yutao Zhu et al. *Large Language Models for Information Retrieval: A Survey*. 2023. arXiv: 2308.07107 [cs.CL] served as a good starting point to all the applications of LLM in information retrieval.

6.1 Query Rewriter

Query rewriting is transformed by LLMs through techniques like prompting methods and fine-tuning. For example, the Query2Doc model uses LLMs to generate relevant passages, expanding the original query and mitigating vocabulary mismatch issues (Liang Wang, Nan Yang, and Furu Wei. *Query2doc: Query Expansion with Large Language Models*. 2023. arXiv: 2303.07678 [cs.IR]).

Another technique, CoT prompting, refines queries in conversational search, aligning more closely with user intent through iterative prompting and contextual understanding (Rolf Jagerman et al. *Query Expansion by Prompting Large Language Models*. 2023. arXiv: 2305.03653 [cs.IR]).

6.2 Retriever

The retriever component benefits from the inherent knowledge of LLMs by leveraging the extensive world knowledge of LLMs to generate search queries and even search data refinement, enhancing the retrieval of contextually appropriate documents (Zhu et al., *Large Language Models for Information Retrieval: A Survey*, op. cit.).

6.3 Reranker

Reranking employs LLMs to rerank documents retrieved by retrievers to return more aligned results. Beyond finetuning LLMs for reranking purposes, prompting relatively large LLMs can be used to improve the ranking returned.

6.4 Reader

The reader component is notably enhanced by LLMs’ ability to understand and interpret complex text structures. Fine-tuning methods allow LLMs to adapt to specific domains, improving their capacity to extract and present relevant information from selected documents. A typical study in this area demonstrates fine-tuning LLMs to generate relevant document content, combining it with queries for more effective information extraction.

6.5 Integration

We drew inspiration from the model presented in Wang-Chiew Tan et al. *Reimagining Retrieval Augmented Language Models for Answering Queries*. 2023. arXiv: 2306.01061 [cs.CL], adapting its framework to our specific context. The POSTTEXT model enriches semi-parametric architectures with three main components:

- **Query Analyzer and Planner (QAP)**: Decomposes complex queries, identifies relevant data sources, and plans efficient execution.
- **Views**: Aggregates data from diverse sources (text, images, videos, tabular data), both publicly and privately sourced.
- **Provenance-Aware Answer Generator (PAG)**: Generates answers with origin tracing and quality assessment.

Inspired by these components, we developed an analogous model:

- **Our QAP**: Implements a Query Rewriter LLM for enhanced query reformulation.
- **Our Views**: Simplifies the original component using an Elasticsearch Node with Google Places API data.
- **Our PAG**: Produces answers from Elasticsearch and cross-verifies with data from a Google Node.

Overall, our model’s architecture and pipeline are heavily inspired by the original framework.

7 Dataset

In constructing our dataset, we examined different data sources before settling on Google Maps. The data sources considered are Web Extracted Data, Yelp API (crowd-sourced reviews about businesses) and Google Maps API. After considering the relevance, availability and recency of the data sources, we concluded Google Maps data would best suit our use case.

7.1 Exploration 1: Web Extracted Data

Processed HTML files were scrapped and used to create an initial data source using the following methods

1. Searching for terms in Google Search and subsequently using Beautiful Soup to crawl and pull-out data from links retrieved in our original Google Search query.
2. Our second method included obtaining sitemap information from popular local (Singaporean) food blogs and subsequently using Beautiful Soup to crawl and pull-out data from links listed in the sitemaps.

7.1.1 Limitations of Web Extracted Data

The web extracted data was not very useful as it contained large amounts of plain text data that was unformatted and hard to process. Without proper labeled data, the dataset would be extremely messy and unorganised to extract useful information from in future steps.

Example:

```
["Filter By(Change?)Going to Don Don Donki (DDDK) is pretty much a
national pastime at this point, especially since the Japanese
supermarket constantly brings in new must-try desserts and snacks
. We recently saw intel that DDDK has brought in a new mochi
sandwich dessert, spotted by Instagram user @breadinosaur.\nImage
credit: @breadinosaurThere are apparently a handful of different
mochi sandwich flavours, including a Strawberry Rare Cheese
flavour reminiscent of the strawberry cream cheese snacks popular
in Japan.\nImage credit: @breadinosaurPriced at around $5, this
mochi sandwich comes with super soft rice mochi and a cream
cheese filling, with a strawberry jelly centre. The dessert is
sold in a pack of two, and is found in DDDKs dessert section.\n
Image credit: @breadinosaurThis mochi sandwich is not to be
confused with the Tokyo Waffle Club waffle mochi sandwiches that
DDDK also sells as a dessert. For these, thin, crispy meringue
waffles encase a cream filling and a flavoured mochi centre.\n
Image credit:\xa0KL FoodieThe waffle mochi snack comes in three
flavours: chocolate, custard, and jam.\nImage credit: @donkisgIn
other Donki news, the supermarket has brought their Wagyu kushi
skewers to Singapore, after the product went viral in Malaysia.
These skewers are available at $4.90, and are made with 100%
Japanese Wagyu beef sourced from the Kagoshima prefecture.The
skewers are exclusively available at 100AM and Orchard
```

Cineleisure outlets. Don Don Donki is also currently running their Great Singapore Sale, till 31 July. Get popular snacks such as their Pretz sticks, flavoured Swiss roll, Pocky, and many more at a discount. Cafe Hachi: Famous Matcha Cafe From Japan Opens In Singapore At Clarke Quay For more Japanese food news, famed Fukuoka matcha cafe, Cafe Hachi, has opened their first outlet outside of Japan, in Clarke Quay. Popular Japanese pon de ring doughnut sellers, Mister Donut, have also opened their first local pop-up stall in Jurong Point! Featured image adapted from @breadinosaurs Get more stories like this. Drop us your email so you won't miss the latest news. Being addicted to caffeine is a personality, right? Eatbook offers honest reviews and recommendations on Singapore food. Eatbook.sg 2023 - All Rights Reserved. \xa0 Eatbook is part of TSL Media Group. Past Contest Winners. \xa0 Privacy Policy"]

Apart from crucial information such as opening hours and address being missing, the above data source also does not perform well with BM25. BM25 puts significant weight on term frequency and the document length, which might cause a long, largely unstructured document to perform badly even if it was relevant. The presence of multiple useful contexts within a single document (specifically in listicles might be less useful for the LLM to take in as context). Hence, a query would likely fail to retrieve favourable documents.

7.2 Exploration 2: Yelp API

Yelp.com publishes crowd-sourced reviews about businesses. We decided to use their API to obtain semi-structured information about businesses using their API. The data provided would be more structured, context-rich as opposed to web-extracted data.

Listed below is an example of Yelp API output:

```
{'query': {'location': 'Changi City Point',
  'term': 'indian',
  'categories': 'restaurants',
  'price': [2, 3],
  'open_now': 'true',
  'attributes': ['parking_lot', 'liked_by_vegetarians'],
  'limit': '20'},
  'results': {'businesses': [{'id': 'WjpaNwaJB9IX7gja2BhBzA',
    'alias': 'tandoori-flame-boon-tat-singapore',
    'name': 'Tandoori Flame @ Boon Tat',
    'image_url': 'https://s3-media3.fl.yelpcdn.com/bphoto/
      wAg83o3_Hgl0ehoG2EvgGA/o.jpg',
    'is_closed': False,
    'url': 'https://www.yelp.com/biz/tandoori-flame-boon-tat-
      singapore?adjust_creative=P2E-4WvnsgTHDYWehmwoCA&
      utm_campaign=yelp_api_v3&utm_medium=api_v3_business_search&
      utm_source=P2E-4WvnsgTHDYWehmwoCA',
```

```

'review_count': 9,
'categories': [{'alias': 'indpak', 'title': 'Indian'}],
'rating': 5.0,
'coordinates': {'latitude': 1.28116774503181,
'longitude': 103.848429529976},
'transactions': [],
'price': '$$',
'location': {'address1': '5 Boon Tat St',
'address2': '#01-01',
'address3': '',
'city': 'Singapore',
'zip_code': '069613',
'country': 'SG',
'state': 'SG',
'display_address': ['5 Boon Tat St',
'#01-01',
'Singapore 069613',
'Singapore']}],
'phone': '+6562233200',
'display_phone': '+65 6223 3200',
'distance': 14002.366642933746,
'reviews': {'reviews': [{'id': 'ta8xkTsYs4CsKc4q--MGhw',
'url': 'https://www.yelp.com/biz/tandoori-flame-boon-tat-singapore?adjust_creative=P2E-4WvnsgTHDYWehmwoCA&hrid=ta8xkTsYs4CsKc4q--MGhw&utm_campaign=yelp_api_v3&utm_medium=api_v3_business_reviews&utm_source=P2E-4WvnsgTHDYWehmwoCA',
'text': "Pro tip: \n- Recommend the Chicken tiki Masla. Skip the Chicken Tandoori.\n\nWhile I'll admit I'm no expert in Indian food, I still feel this places had some...",
'rating': 4,
'time_created': '2020-01-23 05:23:51',
'user': {'id': 'mqxwq4sqdI8lf1iHD2aQRw',
'profile_url': 'https://www.yelp.com/user_details?userid=mqxwq4sqdI8lf1iHD2aQRw',
'image_url': 'https://s3-media4.fl.yelpcdn.com/photo/FelFiza1sIt3zk9NsIUCHA/o.jpg',
'name': 'George P.'}]},

```

7.2.1 Limitations of Yelp API

After extensive testing, we identified that the results from the Yelp API lacked in two major areas:

- **Recency:** The lack of popularity of Yelp, especially in Singapore, meant that there were often incomplete or missing results as users no longer uploaded reviews to Yelp's platform.
- **Robustness:** Apart from not necessarily giving the best results, Yelp did not give some information, such as opening hours and other properties.

This meant that we could not rely on the context provided by Yelp to be the most updated and relevant.

7.3 Chosen: Google Places API

Utilizing Google Places' Text Search and Place Details APIs, we access a rich set of factual information and user reviews about restaurants. To align with typical query requirements, we focused on extracting nine key properties:

1. Name
2. Address
3. Phone Number
4. Availability of Delivery Service
5. Availability of Dine-In Service
6. Opening Hours (categorized by day of the week)
7. Price Level
8. Overall Customer Rating
9. Customer Reviews

Our initial use of the Google Places API was to populate our Elasticsearch database. By inputting over 80 general food-related queries into the API, we compiled a comprehensive list of recommended dining venues in Singapore. This data was then formatted into a JSON file and uploaded to Elasticsearch.

Examples of the queries we used:

```
# in Singapore type queries
"Best restaurants in Singapore",
"Top-rated seafood restaurants in Singapore",
"Vegetarian restaurants in Singapore",
"Restaurants with outdoor seating in in Singapore",
"Vietnamese restaurants in Singapore",
"Middle Eastern cuisine in Singapore",
"Spanish restaurants in Singapore",
"African cuisine in Singapore",

# In Singapore type queries
"Italian restaurants in Singapore",
"Asian cuisine near Singapore",
"Fine dining restaurants in Singapore",
"Brunch places in Singapore",
"Local fast food chains in Singapore",
"Michelin-starred restaurants in Singapore",
```

```
# Locality type queries
"Italian cuisine in Orchard",
"Caribbean dining in Kallang",
"African cuisine in Jurong",
"Scandinavian food in Bukit Merah",
"Uzbek dining in Bukit Panjang",
"Colombian cuisine in Bukit Timah",
"Venezuelan food in Telok Blangah",
"Laotian restaurants in Jurong West"
```

In our secondary application of the API, the Google Places API served as an evaluation tool. Upon receiving a user’s query, we used the API to fetch relevant data, which was then cross-referenced with the Elasticsearch results for accuracy and calculated the precision of our system.

Listed below is the formatted context from Google Places API:

```
{'Option': 1,
'Name': 'Joie | Vegetarian Casual Fine Dining',
'Address': '181 Orchard Rd, #12 - 01, Singapore 238896', 'Phone':
'9663 0901',
'Delivery Available': 'No',
'Dine-In Available': 'Yes',
'Opening Hours':
'Monday: 12:00 3 :00PM, 6:00 10 :00PM\n
Tuesday: 12:00 3 :00PM, 6:00 10 :00PM\n
Wednesday: 12:00 3 :00PM, 6:00 10 :00PM\n
Thursday: 12:00 3 :00PM, 6:00 10 :00PM\n
Friday: 12:00 3 :00PM, 6:00 10 :00PM\n
Saturday: 12:00 3 :00PM, 6:00 10 :00PM\n
Sunday: 12:00 3 :00PM, 6:00 10 :00PM\n',
'Price Level': 3,
'Rating': '4.5',
'Reviews': '\n\nNice vegan and vegetarian option. But the vegan food
was lacking in flavor. The atmosphere is real'}'
```

8 Overall Architecture

In line with the problems discussed in the previous section, we propose *hungry-rag*, a novel retrieval augmented technique to improve the efficiency of LLMs when performing food-related information retrieval. Before going into the finer details of our project, we feel it would be important to provide an overall architecture of our solution to better understand the subsequent sections as shown in Fig. 1.

8.1 Main Pipeline

1. Query Rewriter

- Based on user query this node seeks to prompt user with further clarifying questions about their preferences.
2. Annotate Needs
- Based on clarifying answers from the user, this node seeks to return a list of requirements of attributes that food establishment results must have.
3. Search Term Generator
- This node takes in all output from the previous nodes such as original query, clarifying questions/answers and list of required attributes. The node then outputs search terms that are used to query Elasticsearch or Google Places API(used for evaluation).
4. ESNode (Elasticsearch Information Retrieval Node)
- Elasticsearch Database is built by querying Google Places API using 80+ random food related queries
 - Use search terms from previous node to retrieve output using BM25 algorithm.
5. Reader
- This node processes all outputs produced by previous nodes including the original query, clarifying questions/answers, list of required attributes and context.
 - It then generates the final output that the user views, which contains a list of recommended food establishments.
6. Word Match Citation Generator
- This last node of the main pipeline takes in the final list of recommended food establishments to be returned to user as well as context generated by Elastic Node.
 - If both results match, a citation will be generated.

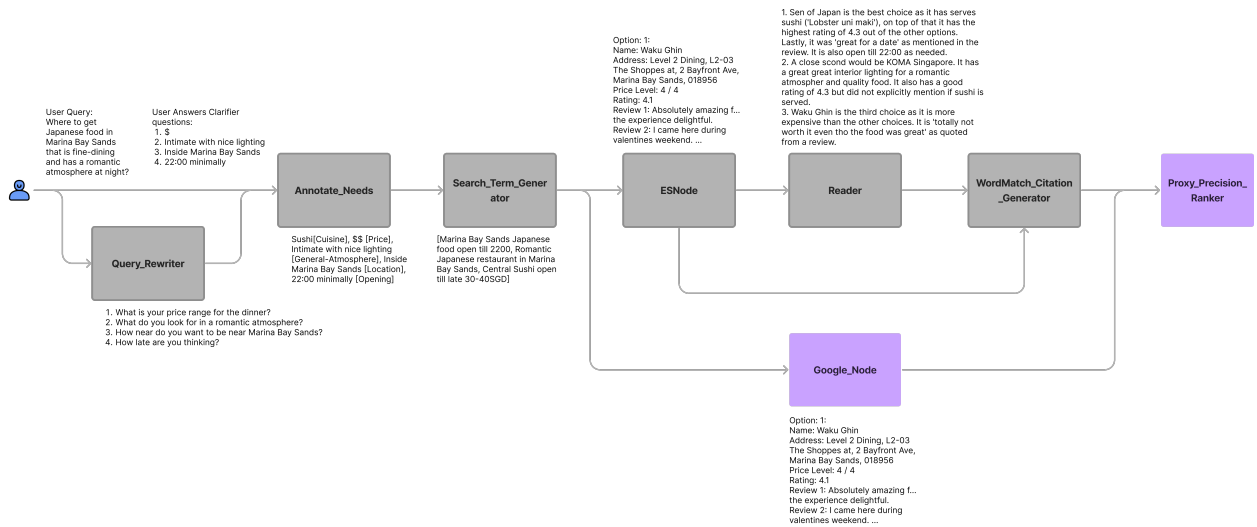


Figure 1: Overall Solution Architecture

8.2 Google Evaluation Pipeline

As Google is often regarded as one of the top search engines, we used their API to compare results. The Google Evaluation pipeline contains additional components to evaluate the main pipeline.

1. Google Node

- Similar to the ESNode, this node takes in search terms and performs a live API call to Google Places API.
- Results returned from the API are used as context that will be compared in the next node.

2. Proxy Precision Ranker

- By comparing context generated from Google Node and the final results generated in the main pipeline. We use proxy precision 20 to produce an evaluation score.

9 Retrieval

In our complete solution, we have made use of an information source that was developed by running 80+ queries on Google APIs as mentioned earlier. This source has been indexed in an Elasticsearch Database. We managed to generate a comprehensive list of recommended restaurants/eateries based on the different requirements we covered in our general query list.

We use Elasticsearch 7.0 API for the purpose of this project. The default scoring algorithm used by Elasticsearch 7.0 is BM25.

Elasticsearch 7.0 harnesses the power of the Okapi BM25 algorithm to measure similarity across indexed documents, facilitating the ranking of retrieved documents. Okapi BM25, a widely acknowledged information retrieval model, is instrumental in evaluating the relevance of documents to a given search query. As users submit search queries, Elasticsearch employs BM25 to scrutinize the indexed documents, taking into consideration key factors such as term frequency, document length, and inverse document frequency.

The BM25 algorithm assigns scores to documents based on their similarity to the user's query, resulting in a ranking that prioritizes documents with the highest degree of relevance.

We query our Elasticsearch Index through the code below:

```
def runQuery(self, infoQuery):
    search = self.es.search(
        index="food_index",
        body={
            "size":3,
            "query":{
                "bool":{
                    "must":[
                        {"match":{"query":infoQuery,
                                }}
                    ]
                }
            }
        })
```

```

    }
  }
)

```

- `"bool":{}` - This indicates that the query is a boolean query, which means that it combines multiple subqueries using logical operators.
- `"must":[]` - This indicates that the subqueries must all match in order for a document to be returned.
- `"match":{"query":infoQuery,}}` - This is a match query that searches for the parameter `infoQuery` in all fields.

10 Generation

We used the LLM Model `vicuna-7b-v1.5` in our project. We employed this model for our project for various reasons, some of which are mentioned below:

1. **High Accuracy:** Demonstrates state-of-the-art accuracy across various natural language processing (NLP) tasks, including question answering, summarization, and text generation. Its performance rivals that of larger language models while maintaining a balance in efficiency and effectiveness.
2. **Efficiency:** Efficient model, requiring minimal computational resources for training and deployment. Its efficiency makes it suitable for resource-constrained environments and real-time applications.
3. **Versatility:** Versatile model, capable of handling a wide range of NLP tasks, including text classification, translation, and dialogue generation. Its versatility makes it a valuable tool for developers and researchers working on diverse NLP projects.
4. **Open-Source:** Open-source model, freely available for anyone to use and modify. Open-source accessibility promotes collaboration, innovation, and wider adoption of the model.

10.1 LLM Usage:

Throughout our project we have 4 different instances where an LLM model was used.

- Writing clarifying questions for ambiguous parts of the query
- Generating search queries
- Annotation of needs
- Generation of conversational-like answers of retrieved result while also reranking relevant restaurant entries as the reader

10.1.1 Query Rewriter

Based on user query this LLM instance seeks to prompt user with further clarifying questions. To get important information a lower temperature of 0.4 is used to ensure more "essential" clarifying questions are asked. For example based on the query "Where to get Japanese food in Marina Bay Sands that is fine-dining and has a romantic atmosphere at night? ", this LLM asks clarifying questions such as:

1. What type of Japanese food are you looking for? e.g. sushi, ramen, teppanyaki, etc.
2. Do you have a preference for specific dishes or would you like to try a variety of options?
3. Are you looking for a specific price range for the food?
4. What are you looking for in a romantic atmosphere?
5. How late at night are you eating?

10.1.2 Annotate Needs

Based on the answers entered by user in the rewritten query, the LLM then generates a list of requirements, for example:

Requirements: Sushi[Cuisine], \$\$ [Price], Intimate with nice lighting
[General-Atmosphere], Inside Marina Bay Sands [Location], 22:00 minimally [Opening]

This is to use LLM's reasoning ability to consolidate and summarise important information about the query into useful user needs for better results.

10.1.3 Search Term Generator

This instance takes in all output from the previous output such as original query, clarifying questions/answers and list of requirements. Using these inputs this LLM instance outputs search terms that are used to query either Google Places API or Elasticsearch. To get slightly more varying and "creative" search terms we increase temperature slightly to 0.8. An example of search results that should be generated is shown below:

Marina Bay Sands Japanese food open till 2200, Romantic Japanese restaurant
in Marina Bay Sands, Central Sushi open till late 30-40SGD

10.1.4 Reader

This last LLM instance processes output previously generated from LLMs as well as additional context from either the Google Places API Node or Elasticsearch Node depending on the pipeline. It essentially acts as a summary that generates the final output that the user views. To get slightly more varying and "creative" search terms we maintain the temperature as 0.8. An example of an answer that should be generated is shown below:

Answer:

1. Sen of Japan is the best choice as it has serves sushi ('Lobster uni maki'), on top of that it has the highest rating of 4.3 out of the other options. Lastly, it was 'great for a date' as mentioned in the review. It is also open till 22:00 as needed.
2. A close second would be KOMA Singapore. It has a great great interior lighting for a

romantic atmosphere and quality food. It also has a good rating of 4.3 but did not explicitly mention if sushi is served.

3. Waku Ghin is the third choice as it is more expensive than the other choices. It is 'totally not worth it even tho the food was great' as quoted from a review.

11 Verification and Citation Generation

The reader generated answer is verified against the context for fact checking. When the reader output is determined to be based on the context, then a citation is generated for that particular context given.

11.1 Creating Annotations for LLM Output Based on Context

We created annotations for the LLM output by assessing how well the output aligns with the given context. This process involves the evaluation of relevance and accuracy between the LLM output and input context. Based on a survey of the existing literature we decided to test the following techniques: BERTScore, word matching, and direct LLM interrogation (asking the LLM yes/no questions) for this purpose. Below, we discuss these techniques and their implementation.

Evaluation Methodology The evaluation involved generating textual outputs based on two types of contexts:

1. Related Contexts: These are contexts used to generate the output and are contextually important
2. Unrelated Contexts: These are contexts not used to generate the output and are not contextually important

11.1.1 BERTScore for Semantic Similarity Assessment

BERTScore is a metric for evaluating the quality of text generated by models based on BERT embeddings. It calculates the cosine similarity between the embeddings of the tokens in the predicted and reference texts.

The following pseudocode describes the process of comparing the BERTScore of generated outputs with a predefined threshold to determine the need for annotation:

Algorithm: BERTScore Comparison

Input: context_list, output_list, threshold

Output: annotation_list

Begin

```
for each context in context_list
    score = calculate_BERTScore(context, output)
    if score < threshold
        annotation_list.add(True)
    else
        annotation_list.add(False)
end for
return annotation_list
```

End

Results The following table presents the BERT scores for each generated output against the respective contexts. See the generated outputs, references used for this benchmarking (Appendix A):

Context Type	Output 1			Output 2			Output 3		
	P	R	F1	P	R	F1	P	R	F1
Relevant									
Option 1	0.7356	0.8179	0.7745	0.7301	0.8114	0.7686	0.7002	0.7984	0.7461
Option 2	0.7408	0.8211	0.7789	0.7308	0.8023	0.7649	0.7358	0.8242	0.7775
Option 3	0.7157	0.7995	0.7553	0.7329	0.8257	0.7765	0.6975	0.8084	0.7489
Irrelevant									
Option 1	0.7819	0.8094	0.7954	0.7738	0.7977	0.7856	0.7757	0.8123	0.7936
Option 2	0.7913	0.8137	0.8023	0.7787	0.7976	0.7880	0.7825	0.8138	0.7979
Option 3	0.8000	0.8144	0.8071	0.7889	0.8023	0.7955	0.7964	0.8278	0.8118

Table 1: Comparison of BERT Scores for Different Options and Context Types

11.1.2 Matching Tokens

This method involves comparing each token in the generated output (specifically, the names of restaurants) with the corresponding context information to determine a match.

The matching process was quantified by counting the number of tokens in the generated output that directly correspond to the tokens in the context. Additionally, the percentage of matching tokens relative to the total length of the generated text is also calculated.

A threshold of 0.1 for the matches to be considered a relevant context as this threshold best divides the relevant and irrelevant contexts for our use case.

Context Type	Output 1			Output 2			Output 3		
	Rest. Name	Match Tokens	% Match	Rest. Name	Match Tokens	% Match	Rest. Name	Match Tokens	% Match
Relevant									
Option 1	FALSE	11	0.1571	TRUE	8	0.1290	FALSE	7	0.1795
Option 2	TRUE	12	0.1714	TRUE	12	0.1935	TRUE	8	0.2051
Option 3	TRUE	6	0.0857	TRUE	10	0.1613	TRUE	6	0.1538
Irrelevant									
Option 1	FALSE	4	0.0571	FALSE	3	0.0484	FALSE	6	0.1538
Option 2	FALSE	3	0.0429	FALSE	2	0.0323	FALSE	3	0.0769
Option 3	FALSE	4	0.0571	FALSE	2	0.0323	FALSE	6	0.1538

Table 2: Comparison of Token Matching in Restaurant Names Across Outputs

11.1.3 LLM-based Judgement

This method prompts an LLM to make the judgement of whether an output is relevant or not. This is done by providing the LLM with the following prompt:

```
prompt = f"""
```

```
In the following task, you are to determine whether the LLM's  
output is relevant to the provided context. Two examples are  
given to illustrate how to assess relevance. Carefully  
compare the LLM's output with the given context, and then  
state whether the output is supported by the context.
```

```
EXAMPLE 1:
```

```
LLM Output:
```

1. Sen of Japan is the best choice as it serves sushi ('Lobster uni maki'), and it has the highest rating of 4.3 out of the other options. Additionally, it was described as 'great for a date' in a review and is open until 22:00.
2. A close second is KOMA Singapore. It has excellent interior lighting for a romantic atmosphere and quality food. It also has a rating of 4.3 but doesn't explicitly mention serving sushi.
3. Waku Ghin is the third choice as it is more expensive than the others. A review stated it was 'totally not worth it even though the food was great'.

```
Context for Example 1:
```

```
Name: KOMA Singapore
```

```
Address: 2 Bayfront Ave, # B1 - 67, Singapore 018972
```

```
Price Level: None / 4
```

```
Rating: 4.3
```

```
Review 1: Really nice restaurant with great food and perfect  
service. The atmosphere can get quite uncomfortable  
because it is always packed. Prices are reasonable for  
quality and location. Very recommended for date nights.
```

```
Review 2: Obviously, the main attraction would be the  
interior lighting. Can't deny the entire place makes a  
great backdrop for IG-worthy photos.
```

```
Review 3: Celebrated my wife's birthday there recently, was  
a very nice place. Food was truly amazing and the staffs  
were very friendly and helpful in many ways.
```

```
Is the LLM Output Supported by the Context for Example 1?
```

```
Answer: Yes, it is relevant.
```

```
EXAMPLE 2:
```

```
LLM Output:
```

1. Sen of Japan is the best choice as it serves sushi ('Lobster uni maki'), and it has the highest rating of 4.3 out of the other options. Additionally, it was described as 'great for a date' in a review and is open until 22:00.

2. A close second is KOMA Singapore. It has excellent interior lighting for a romantic atmosphere and quality food. It also has a rating of 4.3 but doesn't explicitly mention serving sushi.
3. Waku Ghin is the third choice as it is more expensive than the others. A review stated it was 'totally not worth it even though the food was great'.

Context for Example 2:

Name: Saizerya
 Address: 3 Changi City Ave, Singapore 890233
 Price Level: 3 / 4
 Rating: 2.3
 Review 1: Terrible restaurant with bad service, food is bad.
 Review 2: Pasta was delicious and was really cheap. Enjoyed the food.
 Review 3: Had a medium time at Saizerya. It is what you expect.

Is the LLM Output Supported by the Context for Example 2?

Answer: No, it is not relevant.

ACTUAL TASK:

LLM Output:

{context}

Context:

{answer}

Is the LLM Output Supported by the Context in the Actual Task?

Answer:

""

Then check if "is relevant" or "yes" is in the generated text. The outcome is as shown:

Context Type	Option	Output 1	Output 2	Output 3
Relevant	Option 1	TRUE	FALSE	FALSE
	Option 2	FALSE	FALSE	FALSE
	Option 3	FALSE	FALSE	FALSE
Irrelevant	Option 1	FALSE	FALSE	FALSE
	Option 2	FALSE	TRUE	FALSE
	Option 3	FALSE	TRUE	FALSE

Table 3: Comparison of Relevance Across Different Outputs and Options

11.2 Selection

Among the three methods considered, the naive matching token approach seems most effective at distinguishing relevant from irrelevant contexts. Consequently, the final process will employ the percentage of matching tokens and the presence of the restaurant name in the output as criteria to decide whether a context should be marked as a citation.

12 Evaluation

A dataset of a set of queries q shown in Table 4 was passed through Google’s BARD to find out a list of restaurants that satisfied the query. BARD was chosen against other SOTA models like GPT-3.5 due to the quality of responses generated by BARD- which augments its search results by integrating with Google Places API itself. The restaurants were also fact-checked by clicking into the links to see if the results produced by BARD were indeed an accurate match for the query. The output that we get from this stage is ordered in list $l_{i,1}$, corresponding to query q_i . We use list $l_{i,1}$ as the ground truth for that query i .

Further we generated a ranked list of outputs for the same set of queries based on our hungry-rag LLM. Let’s call this ordered list $l_{i,2}$. The objective here is to evaluate the result $l_{i,2}$ against $l_{i,1}$ $\forall q_i$.

Here, we assume that the list $l_{i,1}$ and $l_{i,2}$ consist of only the restaurant names. This is because objectively speaking, the most important consideration of our LLM output are the restaurants, and the respective rank at which they appear. For this reason, we choose to keep only the names of the restaurants for the purpose of evaluation. Although we agree that there might be some information loss caused by discarding the rest of the sentence in the evaluation, we will keep one metric to help us see the pairwise similarity between the $l_{i,1}$ and $l_{i,2}$. Here, we make use of the following metrics:

1. **Jaccard Similarity:** Simply given as the union of sets divided by the intersection.

$$J(l_{i,1}, l_{i,2}) = \frac{|l_{i,1} \cap l_{i,2}|}{|l_{i,1} \cup l_{i,2}|}$$

We may use Jaccard similarity to get a rough sense of how many results are matching from the LLM output. Note that, however, Jaccard similarity does not give us an idea of the position at which the restaurants appear in the result list $l_{i,2}$. In order to get a fair evaluation we truncate the length of the two lists to $\max(\text{length}(l_{i,1}), \text{length}(l_{i,2}))$.

2. **Mean Average Precision:** The number of results obtained by the two queries may not be the same. As such, taking an average precision and recall value helps to normalize by the length of the sets.

$$MAP = \frac{\sum_{i=1}^n \text{Precision}@i}{n} \quad (1)$$

We choose precision here instead of recall for the purpose of evaluation because

1. Answer generation is a *time-critical task*. This means that the number of relevant retrieved documents as a factor of the total retrieved documents is more important than recall, which just considers the number of relevant documents as a factor of the total number of relevant documents.
2. Restaurant finding involves *information redundancy*. There may be many good answers when it comes to finding good restaurants.

	Queries
1	Where to get Japanese food in Marina Bay Sands that is fine-dining and has a romantic atmosphere at night?
2	Where to get Indian food in Singapore that is near the Esplanade and suitable for a group of 10 people?
3	Where to get budget-friendly Mexican food in Orchard Road with vegan options?
4	Where to get halal food in Geylang Serai which is open for breakfast and has a casual setting?
5	Where to get Indian food in Singapore that is near the Esplanade, suitable for a group of 10 people, and has a budget-friendly price tag?
6	Where to get Korean food in Singapore that is near the Botanic Gardens and has a vegetarian menu?
7	Where to get seafood in Sentosa that is family-friendly and has a view of the beach?
8	Where to get Thai food in Clarke Quay that is budget-friendly and has a vegan menu?
9	Where to get Chinese food in Chinatown that is fine-dining and has a romantic atmosphere?
10	Where to get Japanese food in Orchard Road that is near the MRT station and has a gluten-free menu?
11	Where to get Indian food in Little India that is open late and has a view of the street?
12	Where to get Italian food in Robertson Quay that is suitable for a group of 10 people and has a budget-friendly price tag?
13	Where to get Mexican food in Clarke Quay that is family-friendly and has a casual setting?
14	Where to get halal food in Tampines that is open for breakfast and has a vegetarian menu?
15	Where can I find Japanese ramen in Tiong Bahru that is open late at night and suitable for a casual date?
16	Which Thai restaurants in Bugis are known for their family-friendly atmosphere and can accommodate a group of 8 people?
17	What are some popular seafood restaurants in East Coast Park that offer a lively atmosphere and are suitable for a birthday celebration?
18	Where can I get vegetarian-friendly Korean food in Tanjong Pagar with a modern and trendy vibe for a casual lunch with friends?
19	Which French cafes in Robertson Quay provide a quiet and conducive environment for studying or remote work during the daytime?
20	What restaurants in Simei provide Halal Indian Food?

Table 4: List of queries

12.1 Baseline LLM Generation

We made use of the metrics described above to get a sense of the LLMs performance without any retrieval assistance. In order to do this, we considered the following scenarios:

1. **No Context from ES Node + No Additional Criteria (Raw LLM performance):**
The objective of this run was to see if the LLM is able to generate any meaningful results on it's own with no context provided at all.
2. **Bare Context from ES Node + No Additional Criteria:** As mentioned, the data is pre-indexed in Elasticsearch Vector Store. The objective of this run is to see how the LLM performs, if we remove the Additional Criteria provided by the user.

The results of the above approach can be summarized in Table 5.

	Scenario 1	Scenario 2
MAP(Average \forall i)	0.0498	0.1786

Table 5: Mean Average precision for baseline LLM

Deductively, we can see a 4-fold improvement in precision by just adding the additional context from the Elasticsearch node.

In summarizing the evaluation of the hungry-rag pipeline against the ground truth list l_1 , using Jaccard Similarity Score and MAP, we observe key findings. The full pipeline's precision score is lower when compared to the version without additional criteria. This variance arises because the State-of-the-Art (SOTA) generated answers do not incorporate additional criteria like hungry-RAG, which uses these criteria to construct search queries for Elasticsearch. However, the accuracy of reflecting the original user query in search terms varies.

For example, the query "What restaurants in Simei provide Halal Indian Food?" resulted in search terms like "Tandoori Restaurants in Simei" and "Halal Indian Food in Simei". These terms might not fully capture the query's intent, potentially leading to information loss. In contrast, a query like "Casual fine dining restaurant with sushi in Marina Bay" effectively captured relevant search terms, resulting in higher Jaccard and MAP scores of 0.5 and 0.386, respectively.

A proposed solution involves combining contexts from all search terms of a query. However, as noted by Nelson et. al, 2023, this could lead to long input contexts, posing challenges in extracting relevant information, particularly if it doesn't appear near the beginning or end. Therefore, while combining search terms may offer a broader context, it doesn't guarantee improved answer quality due to potential difficulties in managing long contexts.

The expressiveness of the hungry-rag pipeline also rivals the SOTA generated answers with its ability to explain the reasoning for the choice of restaurant.

12.2 Analysis against SOTA-Generated Answers

In this section, we will discuss the results of the evaluation of the full hungry-rag pipeline against the ground truth list l_2 . We will be making use of both Jaccard Similarity Score as well as MAP to

get a sense of the model’s performance. The results have been summarized in table 5. Notably, the precision score of the full pipeline is lower than that of the pipeline without the additional criteria.

Here, there are two things happening. First, the SOTA generated answers do not make use of additional criteria. Note that hungry-RAG constructs and passes in search queries into Elasticsearch, based on additional criteria. The result generated is the result of the first query that generates a context that is not empty. We presume that each search query passed into ES is a variation of the combination of additional criteria input from the user. In an ideal scenario, there should be no loss of information in the search queries passed. But we note that this is not the case for all the queries. Consider the reconstruction of additional criteria in query 20 from Table 4:

What restaurants in Simei provide Halal Indian Food?

The search terms generated in the intermediate steps are as follows:

Tandoori Restaurants in Simei, Halal Indian Food in Simei, Cozy Restaurants in Simei, Indian Cuisine in Simei, 'All Spices in Indian Food in Simei, Simei Location

We can see that the search terms generated here are not a complete reflection of the original query passed in. The first query would return Tandoori restaurants in Simei, and the second query would return Halal Indian food in Simei. There is a possibility that the two sets could be mutually exclusive of each other. We must also note here that this is not the case with all the queries. For example query 1 from table 4 generated the search terms given below:

Casual fine dining restaurant with sushi in marina bay, Japanese restaurant in marina bay for a casual date

As the above query conforms to our ideal scenario of capturing all the relevant search terms, we see that the answer generated has a higher Jaccard and MAP score of 0.5 and 0.386 respectively.

Second, a possible solution to this would involve retrieving all contexts for all the search terms generated for a given query and concatenating the results. But, this may lead to a different problem that has been discussed in Nelson F. Liu et al. *Lost in the Middle: How Language Models Use Long Contexts*. 2023. arXiv: 2307.03172 [cs.CL]. The struggle to access and use information in long input contexts is observed in tasks like multi-document question answering and key-value retrieval. It is also shown that performance is often highest when relevant information occurs at the beginning or end of the input context. Our pipeline would indeed produce a larger context if more search terms are passed into Elasticsearch - and the answer generated is likely to be worse on average as the relevant restaurants are less likely to appear near the beginning or the end of a large context if the contexts were concatenated.

	Jaccard Score	MAP
Average	0.1025	0.1596

Table 6: Average Jaccard and MAP scores for full pipeline

12.3 Precision Analysis based on Proxy Ground Truth

This is the evaluation of our Restaurant Answer Generator (RAG) system based on Mean Average Precision (MAP) but using the Google retrieved restaurants as a proxy. Initially, each user query was processed through our custom-built pipeline. The generated response was then compared with a set of documents retrieved from Google’s node, which we designated as our proxy for the ground truth.

This ground truth proxy is based on the top documents (limited to a maximum of 10) retrieved from Google Maps searches. These documents primarily consist of restaurant listings and reviews that are deemed relevant to the query. The underlying assumption for using Google’s search results as a proxy is twofold: firstly, Google’s retrieval strategy is renowned for its efficiency and relevance; secondly, the alternative ground truths are difficult to obtain.

To quantify the system’s effectiveness, we calculated the Mean Average Proxy Precision using the following method:

$$AverageProxyPrecision = \frac{\sum_{i=1}^n Precision@Rank(i)}{n} \quad (2)$$

$$MeanAverageProxyPrecision = \frac{\sum_{j=1}^Q ProxyPrecision(j)}{Q} \quad (3)$$

Here, ‘n’ represents the number of ranks or positions in the generated output, and ‘Precision at Rank i’ is the precision calculated at each individual rank of the search results generated by our Language Learning Model (LLM). The average of these precision provides us with the proxy precision of the model for a specific query. ‘Q’ represents all the queries.

Results show a low mean average proxy precision score of 0.0175. Most of the precision scores are zero. This was expected as the finite scope of indexed results within our Elastic Search database constrained our RAG system. Despite leveraging Google Maps data, our results differed from those obtained directly via the Google Maps API. This discrepancy can be attributed to the inherent differences in the retrieval methodologies. Our system employs the BM25 algorithm, a widely used information retrieval function based on the probabilistic retrieval framework. While effective, BM25 may not match the sophistication of Google’s proprietary algorithms.

With that said our analysis also has a major limitation: the documents only tests the precision of the retrieved results and not the recall. Even if all retrieved restaurants are relevant to query, there might be other relevant restaurants not retrieved. Our assumption that Google Map’s search is a good proxy of ground truth might also be flawed. A more robust approach could involve human evaluators who can manually generate optimal answers. This would help in establishing a more accurate and reliable ground truth.

12.4 Evaluation Repository

The repository for our evaluation sheets for this project can be accessed by clicking this **link**.

12.5 Suggested Improvements

Future enhancements to our RAG system should consider the following aspects:

1. **Data Source Expansion:** The current limitation of data from Elastic Search could be addressed by integrating directly with Google's node. This would ensure access to the most updated and comprehensive data sets, thereby enhancing the relevance and accuracy of the generated responses.
2. **Algorithmic Advancements:** Exploring more advanced retrieval algorithms beyond BM25 could potentially improve the precision of the search results that align with human preferences. Incorporating machine learning techniques and more dynamic ranking algorithms could provide a significant boost to the system's performance.

References

- Jagerman, Rolf et al. *Query Expansion by Prompting Large Language Models*. 2023. arXiv: 2305.03653 [cs.IR].
- Liu, Nelson F. et al. *Lost in the Middle: How Language Models Use Long Contexts*. 2023. arXiv: 2307.03172 [cs.CL].
- Tan, Wang-Chiew et al. *Reimagining Retrieval Augmented Language Models for Answering Queries*. 2023. arXiv: 2306.01061 [cs.CL].
- Wang, Liang, Nan Yang, and Furu Wei. *Query2doc: Query Expansion with Large Language Models*. 2023. arXiv: 2303.07678 [cs.IR].
- Zhu, Yutao et al. *Large Language Models for Information Retrieval: A Survey*. 2023. arXiv: 2308.07107 [cs.CL].

A References for Generated Outputs

Each generated output was compared against a set of relevant and irrelevant references. Below are the details for each:

A.1 Entry 1

Question:

Where to get Indian food in Singapore that is near the Esplanade and suitable for a group of 10 people?

LLM Output:

1. A1 Indian Curry is the best choice as it has North Indian cuisine (which is required) and is open till 7:15 PM, which is a requirement as well. However, the reviews were mixed and one of them mentioned that the quality of one of the curries tasted off, so it is a bit risky to try.
2. Tayyiba Briyani Restaurant @ Nexus One North is a good second choice as it has 4.7 rating and the reviews suggest that it has the best chicken biryani in Singapore. It is also open till 8 PM.\n
3. Pangat North Indian Restaurant is the last choice as it has 3 rating and the reviews are not available. However, it is open till 7 PM, which is also a requirement.\n\nRank: 1 (A1 Indian Curry) > 2

Relevant References:

```
{
  'Option': 1,
  'Name': 'A1 Indian Curry North Indian & Lebanese Cuisine',
  'Address': '12/13 Pahang St, Singapore 198613',
  'Phone': '9423 8377',
  'Delivery Available': 'No',
  'Dine-In Available': 'Yes',
  'Opening Hours': 'Monday: 8:00\u202fAM\u202f\u2020\u2020\u2020 \u2020\u2020\u202012:15\u202f
    \u202fAM\u202f\nTuesday: 8:00\u202fAM\u202f\u2020\u2020\u2020 \u2020\u2020\u202012:15\u202f
    \u202fAM\u202f\nWednesday: 8:00\u202fAM\u202f\u2020\u2020\u2020 \u2020\u2020\u202012:15\u202f
    \u202fAM\u202f\nThursday: 8:00\u202fAM\u202f\u2020\u2020\u2020 \u2020\u2020\u202011:30\u202f
    \u202fPM\u202f\nFriday: 8:00\u202fAM\u202f\u2020\u2020\u2020 \u2020\u2020\u202012:30\u202fAM\u202f
    \u202f\nSaturday: 8:00\u202fAM\u202f\u2020\u2020\u2020 \u2020\u2020\u202012:30\u202fAM\u202f
    \u202f\nSunday: 8:00\u202fAM\u202f\u2020\u2020\u2020 \u2020\u2020\u202012:15\u202fAM\u202f',
  'Price Level': None,
  'Rating': '3.9',
  'Reviews': ['Service was ok..... we enjoyed the Fish and
    Chicken sets but didn\u202f\u202f\u202f like 1 of the curries included
    as it tasted a little off...like it was left in the open
    for too long ...."basi...stale" and we have tasted many
    curries in the past and non-tasted ever like that.....\n\u202f
```

```

        nDidn\'t wait too long for our food..... the gentleman
        who took our order.....he made sure we understood that
        the food we ordered will not be the same as in the pics
        in the menu .... and it sure wasn\'t anything like the
        pics .....\\n\\nOverall we enjoyed our dinner.....just
        minus that dish....',
    'Indian food. We ordered chicken dum biryani, it was very
    disappointing. Chicken portion extremely small and dry.\\n
    I also ordered some falafel which was worse than the
    biryani.\\nThe 1st photo below is what was advertised..
    the 2nd photo is what was presented!!',
    'waited for 30mins for food and still not served yet until i
    asked, and they suddenly served us food that was cold
    and tasteless. will not even come back or recommend this
    place.']*
},
{
    'Option': 2,
    'Name': 'Tayyiba Briyani Restaurant @ Nexus One North',
    'Address': '1 Fusionopolis Link, #01-03 Nexus @one-north -
    Space for Rent, Singapore 138542',
    'Phone': '6993 9648',
    'Delivery Available': 'Yes',
    'Dine-In Available': 'Yes',
    'Opening Hours': 'Monday: 10:30\\u202fAM\\u 2 0 0 9 \\u20098:00\\
    u202fPM\\nTuesday: 10:30\\u202fAM\\u 2 0 0 9 \\u20098:00\\
    u202fPM\\nWednesday: 10:30\\u202fAM\\u 2 0 0 9 \\u20098:00\\
    u202fPM\\nThursday: 10:30\\u202fAM\\u 2 0 0 9 \\u20098:00\\
    u202fPM\\nFriday: 10:30\\u202fAM\\u 2 0 0 9 \\u20098:00\\u202fPM
    \\nSaturday: 11:30\\u202fAM\\u 2 0 0 9 \\u20095:30\\u202fPM\\
    nSunday: Closed',
    'Price Level': None,
    'Rating': '4.7',
    'Reviews': ['Had the best chicken biryani in Singapore here.
    Awesome side dishes and raita. Highly recommended',
    'Excellent service... and yummilicious briyani.\\nMust try!\\n
    Ambience can certainly be enhanced to create an overall
    awesome dining experience.',
    'Amazing biryani paired with tender chicken that will
    literally fall off the bone! Whenever I have a biryani
    craving, I will always think of this restaurant. Owners
    are warm and friendly, providing great service. Highly
    recommended for anyone working around metropolis/
    fusionopolis to drop by!']
},
{
    'Option': 3,
    'Name': 'Pangat North Indian Restaurant',

```



```

        nSaturday: 11:00\u202fAM\u2009 \u20099:00\u202fPM\
        nSunday: 11:00\u202fAM\u2009 \u20099:00\u202fPM",
    "Price Level":2,
    "Rating":"4.1",
    "Reviews":[
        "Ordered the soy sauce garlic chicken drum sticks, the
          seaweed fries and the yuzu drink. The chicken has a
          rich soy sauce taste, is pretty tender inside, has
          crispy skin and is not spicy at all. I believe anything
          with one chilli beside it is not spicy at all. Would
          recommend the chicken and the fries.\n\nOne thing to
          note is that the chicken drum sticks are relatively big
          . I ordered 6 for 2 people thinking that it was small
          but it went from a snack to a full meal.",
        "Place is nice and clean. service is really minor, but
          this is fast food style. But a bit hard for new
          customer. But food. Really good. Very worth the dollar
          .",
        "Got the drumsticks mix - preferred the yangnyeom sauce to
          the soy as it was much tastier. The chicken was
          lightly seasoned, and the batter was crispy, but I have
          had more tender meat elsewhere.\n\nThe fries were
          underwhelming, so I do not recommend getting it as
          sides."
    ]
},
{
    "Option":2,
    "Name":"Chir Chir Fusion Chicken Factory Somerset",
    "Address":"313 Orchard Rd, B3 - 04 05 / 06, Singapore
        238895",
    "Phone":"6509 8364",
    "Delivery Available":"Yes",
    "Dine-In Available":"Yes",
    "Opening Hours":"Monday: 11:00\u202fAM\u2009 \u20099:30\
        \u202fPM\nTuesday: 11:00\u202fAM\u2009 \u20099:30\u202fPM\
        \nWednesday: 11:00\u202fAM\u2009 \u20099:30\u202fPM\
        nThursday: 11:00\u202fAM\u2009 \u20099:30\u202fPM\
        nFriday: 11:00\u202fAM\u2009 \u200910:00\u202fPM\
        nSaturday: 11:00\u202fAM\u2009 \u200910:00\u202fPM\
        nSunday: 11:00\u202fAM\u2009 \u20099:30\u202fPM",
    "Price Level":"None",
    "Rating":"3.8",
    "Reviews":[
        "Visit Date 20 Nov 23\n\nChir Chir is a mandatory lunch
          spot when kids are visiting Singapore. When they were
          in school it used to be more often but now it is like
          once a year affair.\n\nVisited the 313 outlet after 1.5

```


years. Not sure exactly but couldn't locate the Bugis outlet in google. Perhaps it is closed down.\n\nToday it was not that satisfying experience I had expected, visiting after so long. We were seated quickly and the food came fast though.\n\nThe dishes we had\n- Garlicky Soy chicken. There was the familiar sweetness and soy although it tasted a little less garlicky than it used to be! However with that, the chewy toppoki and crunchy sweet potato was great.\nLast time what made this chicken dish heavenly was a small bowl of spicy sauce on the side. This time they said, it was not for sale separately. Fair enough but the experience was not the same without it\n- Chir Cheese Fries Tower. The platter was beautifully presented with the inviting melted cheese on the side\n- Ice lemon tea was very good\n\nChir chir has interesting drinks. Amongst others ...\nPassion Yellow Tok! Tok! A pale yellow Jug of mocktail. Today it was not available.\nBlue lemonade: By the time the blue jug with greenish and yellow hue came on the adjacent table we had ordered our ice lemon tea. but it was very tempting\n\nIt was a lot of food already but i still missed the Topokki and Cheese. It is generally delicious\n\nGuess once a year is good enough",

"Google shown the closing hour is 9.30pm. When we reached at 8pm we were told that the restaurant is closing at 9 pm. If there is a change in the operating hour, it should be updated. We made our way to Somerset only to find out that we have only less than an hour to dine. We felt that staff is rushing us to quickly finish our meal. Overall, we didnt enjoy the whole experience. Chir Chir used to be one of my go to place for korean fried chicken fix but after today experience, i will not want to go back. A few of the dishes that we would like to order was unavailable. Not sure if it is because we reached at 8pm or it is always unavailable. The honey butter chicken is the worst dish ever!!! I DO NOT RECOMMEND The Honey Butter Chicken. It used to be quite tasty with real honey and butter now it is replaced with some honey flavoured powder which smell and taste AWFUL. The spicy toppoki is not bad and the garlicky chicken with the cheese dip is not bad. Price wise its too pricey given that the quality of food is below average.\nOverall experience, sucks! Definitely not going back.",

"Crispy chicken with mustard and salsa - 4 stars\nGood flavour, not sweet. Juicy with a crisp batter. The salsa that came with it was my favourite, but they gave

```

        a tiny portion. The server (with gelled back hair)
        walked past with a gigantic tub of salsa so I asked if
        I could have more. He said it only came with one
        portion. For $35 for the plate of tenders this seems
        rather ridiculous. I offered to pay for it and he still
        refused. Seems like really poor company policy if it
        really is. Docked stars for the poor service."
    ]
},
{
    "Option":3,
    "Name":"JINJJA Chicken Diner @ Bugis Plus",
    "Address":"201 Victoria St, #01-11 Bugis+, Singapore 188067",
    "Phone":"None",
    "Delivery Available":"Yes",
    "Dine-In Available":"Yes",
    "Opening Hours":"Monday: 11:00\u202fAM\u202f\u2009 \u2009\u2009\u2009\u202fPM\u202f
        \u202fPM\u202fTuesday: 11:00\u202fAM\u202f\u2009 \u2009\u2009\u2009\u202fPM\u202f
        \u202fPM\u202fWednesday: 11:00\u202fAM\u202f\u2009 \u2009\u2009\u2009\u202fPM\u202f
        \u202fPM\u202fThursday: 11:00\u202fAM\u202f\u2009 \u2009\u2009\u2009\u202fPM\u202f
        \u202fPM\u202fFriday: 11:00\u202fAM\u202f\u2009 \u2009\u2009\u2009\u202fPM\u202f
        \u202fPM\u202fSaturday: 11:00\u202fAM\u202f\u2009 \u2009\u2009\u2009\u202fPM\u202f
        \u202fPM\u202fSunday: 11:00\u202fAM\u202f\u2009 \u2009\u2009\u2009\u202fPM\u202f
        \u202fPM",
    "Price Level":2,
    "Rating":"4.2",
    "Reviews":[
        "Located on the first floor of Bugis Plus, they serve
        Korean-style fried chicken with flavorful sauce.
        Unfortunately, the chicken wasn't warm enough. The rice
        was incredibly soft! The staff provided good service,
        but the cleaner frequently attempted to clean the table
        even when we hadn't finished our food, which was quite
        annoying. The salad was good and came in a generous
        portion. Sometimes, they offer promotions for the
        chicken. Overall, it was a good experience.",
        "The food taste awefully cold and hard. The fries,
        chicken omg taste so cold and hard. My kids rejected
        them and we have to eat the food worth $50 ourselves!!
        The rice is so cold too not even warm. The 3 coke we
        had all almost no gas, it must been prepared way too
        earlier. The skin suppose to be crispy but its very
        hard and cold. Its the worst jinjja experience",
        "A big thumbs up for this r e s t a u r a n t \nI was
        surprised that it literally had the best first
        impressions.\nEvery menu is getting better every time I
        taste. Not that much has shown the authenticity of
        Korean food. But still the best food to try out when
        y o u r e in Singapore."
    ]
}

```

```
]
}
```

A.3 Entry 3

Question:

Where to get Italian food in Little Italy that is family-friendly and has a casual setting?

LLM Output:

1. Cicheti is the best choice as it has a great atmosphere and it is family-friendly. It also has great reviews for Italian food and has a variety of dishes at an affordable price. It is located in Little Italy and has lunch and dinner service.
2. Positano Risto is a great option as it has good ratings and reviews, it is family-friendly and offers Italian cuisine. It is located in Bussorah</s>

Relevant References:

```
{
  "Option":1,
  "Name":"Little Italy - Katong",
  "Address":"297 Tanjong Katong Rd, Singapore 437080",
  "Phone":"8733 9903",
  "Delivery Available":"Yes",
  "Dine-In Available":"Yes",
  "Opening Hours":"Monday: Closed\nTuesday: 12:00\ u 2 0 0 9 \
    u20092:45\u202fPM, 5:00\ u 2 0 0 9 \u200910:30\u202fPM\
    nWednesday: 12:00\ u 2 0 0 9 \u20092:45\u202fPM, 5:00\ u 2 0 0 9
    \u200910:30\u202fPM\nThursday: 12:00\ u 2 0 0 9 \u20092:45\
    u202fPM, 5:00\ u 2 0 0 9 \u200910:30\u202fPM\nFriday: 12:00\
    u 2 0 0 9 \u20092:45\u202fPM, 5:00\ u 2 0 0 9 \u200910:30\
    u202fPM\nSaturday: 12:00\ u 2 0 0 9 \u20092:45\u202fPM, 5:00\
    u 2 0 0 9 \u200910:30\u202fPM\nSunday: 12:00\ u 2 0 0 9 \u20092
    :45\u202fPM, 5:00\ u 2 0 0 9 \u200910:30\u202fPM",
  "Price Level":2,
  "Rating":"4.3",
  "Reviews":[
    "An impromptu visit turned into an unexpectedly fine meal.\n
    \n\nThe burrata was delicious, served with fresh
    ingredients including sweet tomatoes. The complementary
    bread blew my mind as it was freshly baked, fluffy and
    tasty! Went really well with the burrata. I could've
    just have the bread and burrata as a meal alone!\nFour-
    cheese pizza was good. Nicely baked thin crust.\nStaff
    was friendly. Interior/decorations were bluish simple in
    the good way.\n\n\nAn enjoyable Sunday brunch it was!",
```

```

    "Very nice and attentive service.\nAntipasti: Parma e
      Burrata and Gamberoni in Padella is superb nice!\nPasta:
      Spaghetti Alle Vongole taste is recommended\nDessert:
      Lava Cake is ok but not the chocolate flowy type i
      expected...\nDefinitely recommeded and will again...\n
      nCant help it... have to order the Pizza takeaway for
      supper.....",
    "Visited the tanjong katong outlet a few months back.
      Waiters were hospitable. Food was surprisingly good, too
      , considering the fact that it was lunchtime and the
      restaurant was empty.\n\nI got the bufalina pizza and a
      lasagne. Pizza wasn't the traditional neapolitan style,
      but instead, it was thin and really crispy unlike most
      other Italian spots. Undercarriage char marks were a
      good indication of a solid high-temp oven. Pomodoro was
      fresh and tangy, coupled with proper fresh whole-milk
      mozzarella. You could also taste the same quality
      pomodoro in the lasagne, albeit my only complaint was
      that it felt a little dense.\n\nOverall, good portions
      and nice ambience. P.s. Do try their pizzas if you're a
      fan of thin crust."
  ]
},
{
  "Option":2,
  "Name":"Positano Risto",
  "Address":"66 Bussorah St, Singapore 199479",
  "Phone":"6292 1866",
  "Delivery Available":"Yes",
  "Dine-In Available":"Yes",
  "Opening Hours":"Monday: 12:00\ u 2 0 0 9 \u200910:00\u202fPM\
    nTuesday: 12:00\ u 2 0 0 9 \u200910:00\u202fPM\nWednesday:
    12:00\ u 2 0 0 9 \u200910:00\u202fPM\nThursday: 12:00\ u 2 0 0 9
    \u200910:00\u202fPM\nFriday: 12:00\ u 2 0 0 9 \u200910:00\
    u202fPM\nSaturday: 12:00\ u 2 0 0 9 \u200910:00\u202fPM\
    nSunday: 12:00\ u 2 0 0 9 \u200910:00\u202fPM",
  "Price Level":2,
  "Rating":"4.3",
  "Reviews":[
    "Service is good with friendly staff. Food is served quite
      fast, they are tasty, but all tend to be on the salty
      side.\n\nThe call out for us is the squid ink seafood
      spaghetti and the five cheeses pizza. The thin crust was
      impressive. The spaghetti cooked Al Dante.\n\nThe
      lobster bisque was nice but too salty; the soft shell
      crab for the linguine was not fresh, fried in thick
      batter and tasted weird.\n\nThe ambience is at best so
      s o ",

```

"Went there for a gathering and the acoustics was not made for large gatherings. I do like the salad though the waiter did not ask if I wanted olives in my salad. Service was prompt for a bustling place. The ambience around the restaurant is nice as kg glam has a nice vibe . Nice area to walk around after the meal.",

"Great place for western halal food. We got a few dishes for sharing as we were spoilt for choice.\n\nService was personable and staff was upfront that they had sold out of some items, that we should order dessert in advance and that we would have to wait, which I appreciated. They were also very attentive but also let us sit and not chase us away when we finished our meal.\n\nFood was delicious and I ll be back to try more dishes for sure."

]

},

{

"Option":3,

"Name":"Cicheti",

"Address":"52 Kandahar St, Singapore 198901",

"Phone":"9725 6255",

"Delivery Available":"Yes",

"Dine-In Available":"Yes",

"Opening Hours":"Monday: 12:00\ u 2 0 0 9 \u20092:30\u202fPM, 6:00\ u 2 0 0 9 \u20099:30\u202fPM\nTuesday: 12:00\ u 2 0 0 9 \u20092:30\u202fPM, 6:00\ u 2 0 0 9 \u20099:30\u202fPM\nWednesday: 12:00\ u 2 0 0 9 \u20092:30\u202fPM, 6:00\ u 2 0 0 9 \u20099:30\u202fPM\nThursday: 12:00\ u 2 0 0 9 \u20092:30\u202fPM, 6:00\ u 2 0 0 9 \u20099:30\u202fPM\nFriday: 12:00\ u 2 0 0 9 \u20092:30\u202fPM, 6:00\ u 2 0 0 9 \u200910:00\u202fPM\nSaturday: 12:00\ u 2 0 0 9 \u20092:30\u202fPM, 6:00\ u 2 0 0 9 \u200910:00\u202fPM\nSunday: 12:00\ u 2 0 0 9 \u20092:30\u202fPM, 6:00\ u 2 0 0 9 \u20099:30\u202fPM",

"Price Level":3,

"Rating":"4.3",

"Reviews":[

"was told that it was the best pizza in Singapore... it was pretty good, but not the best. the open concept kitchen is pretty cool, where you can watch the kitchen staff prep food. the interior was very noisy though, and the tables were quite cramped. you can't really fit two pizzas on a table for two with room for drinks.\n\nthe menu was predominantly very meaty, and not very many lighter vegetable options. prices were reasonable.\n\njust FYI the toilet has a slatted door that can be semi-seen through. soooo if you're highly private then ya

",

```

    "The Pork Thermohawk is sweet, juicy and strong aroma of
      char.\nThe mussels with the fresh tomatoes.. sweet, sour
      .\nOnly advice do not seat near the door during the
      afternoon the Heat gets in and was sweating in addition
      to the guests coming in and out , quite horrid to be
      honest, other than that all good      ",
    "Came here for lunch one day as my friend recommended this
      place and was not disappointed.\n\nMust order would
      definitely be the pizzas- doughy with a good amount of
      chewiness but not too thick. Calamari was also fried to
      perfection without being too oily. Crab meat pasta was
      ok although I felt the flavours didnt really infuse
      well into the pasta. The molten chocolate cake was also
      just so so I definitely have had better.\n\nOverall a
      good place for a special lunch/ dinner!"
  ]
}

```

A.4 Irrelevant Context

Irrelevant References used for each generated output:

```

{
  "Option": 1,
  "Name": "Waku Ghin",
  "Address": "Level 2 Dining, L2-03 The Shoppes at, 2 Bayfront
    Ave, Marina Bay Sands, 018956",
  "Price Level": "4 / 4 (0- Free, 1- Inexpensive, 2- Moderate,
    3- Expensive, 4- Very Expensive, None- Price Data Not
    Available)",
  "Rating": 4.1,
  "Reviews": [
    "Absolutely amazing food; had the Omakase at the chefs
      table. The umami and creamy sea urchin balanced with the
      sweetness of the shrimp and slight saltiness of the
      caviar to produce a full flavour profile with each
      mouthful. Loved the perfectly (teppanyaki) seared abalone
      with risoni and the A5 Kobe wagyu dipped in the egg yolk
      (it was so silky, tender and flavourful from the
      marbling). Service was excellent as well; they were
      jovial and attentive and made the experience delightful
      .",
    "This is a late review. I came here during valentines
      weekend. Both my friend and I had food poisoning and I
      was bed ridden for three days. Only decided to review now
      because my friend told me that her group of friends also
      got food poisoning from the same group of restaurants.
      So totally not worth it even tho the food was great.

```

```

        Spent 750 per pax to get food poisoning.. really
        ridiculous..",
    "We had some Izakaya food and beer before going for show.
    Food must be good considering the price is not cheap for
    beer food, but its a tad a bit too salty for me. Service
    is great of course. My favorite will be the lobster
    spaghetti, a bit of fusion."
]
},
{
    "Option": 2,
    "Name": "Sen of Japan",
    "Address": "2 Bayfront Ave, #01-86, Singapore 018972",
    "Price Level": "3 / 4 (0- Free, 1- Inexpensive, 2- Moderate,
        3- Expensive, 4- Very Expensive, None- Price Data Not
        Available)",
    "Rating": 4.4,
    "Reviews": [
        "Lunch with a view. The marugo pizza is really nice and thin
        . Pork jowl is thinly slice. The thinnest I ve seen.
        Nice in paring with the sauce. Black Cod fish is soft and
        nicely cook too. V worth it if you have American Express
        card, to enjoy 50% disc on the bill. This meal of 3 fish
        cause only $80+.",
        "Five stars for everything (except maybe the receptionist
        that received us.. kind of rude). The chefs were super
        friendly (sat at the counter), and I had nice small talk
        with them while eating.",
        "Excellent upscale Japanese restaurant great for a date.
        Lobster uni maki was out of this world. Amazing flavours
        and textures in a single mouthful."
    ]
},
{
    "Option": 3,
    "Name": "KOMA Singapore",
    "Address": "2 Bayfront Ave, #B1-67, Singapore 018972",
    "Price Level": "None / 4 (0- Free, 1- Inexpensive, 2-
        Moderate, 3- Expensive, 4- Very Expensive, None- Price
        Data Not Available)",
    "Rating": 4.3,
    "Reviews": [
        "Really nice restaurant with great food and perfect service.
        The atmosphere can get quite uncomfortable because it is
        always packed. Prices are reasonable for quality and
        location. Very recommended for date nights.",
        "Obviously, the main attraction would be the interior
        lighting. Can't deny the entire place makes a great

```

```
        backdrop for IG-worthy photos.",  
    "Celebrated my wife birthday there recently, was a very nice  
      place. Food was truly amazing and the staffs were very  
      friendly and helpful in many ways."  
    ]  
}
```

B Proxy Precision

This is the result of the proxy precision calculation in:

Evaluation: Google Sheet