Project 2: Prototype IR System

ESAN RESTAURANT SEARCH

By

6488035 Teerut Krachangpoj

6488123 Kitiphum Mueangthongkham

6488125 Phutthikanj Kitivoranondh

6488162 Thachchai Pattamasrirattana

ITCS 414 Information Retrieval and Storage

Semester 1 / 2023

Faculty of Information and Communication Technology

Mahidol University

1.Introduction

The Esan restaurant search system is a platform with numerous features created to satisfy the demands of the food industry and users. It lists numerous Esan restaurants and includes essential data like names, addresses, phone numbers, and reviews. For those looking for Northeastern Thai cuisine, the system aims to provide frequent updates and acts as an entry point for exploration and inspiration. It is a useful tool for both experienced food enthusiasts and people learning about Thai flavors.

Similar systems for esan restaurant search

-Google Maps: Users can search for specific types of Esan food and read reviews, view photos, and find locations on the map.

-Wongnai : User can search for type of cuisine, see review, prices of food, location, opening hour, and photo of restaurant and food.

-TripAdvisor: Same as Wongnai that user can search for type of cuisine that they want, location, and opening hour

-GrabFood : User can search type of food and can order online also can see review and opening hour.

2.Implementation

i. Data collection, example documents, and data statistics

We have gathered our data from a variety of sources, categorized into three main types:

1.Restaurant Review platforms and websites

-Wongnai

-Michelin guide

2.Social Media Platforms

-Instagram

-Facebook

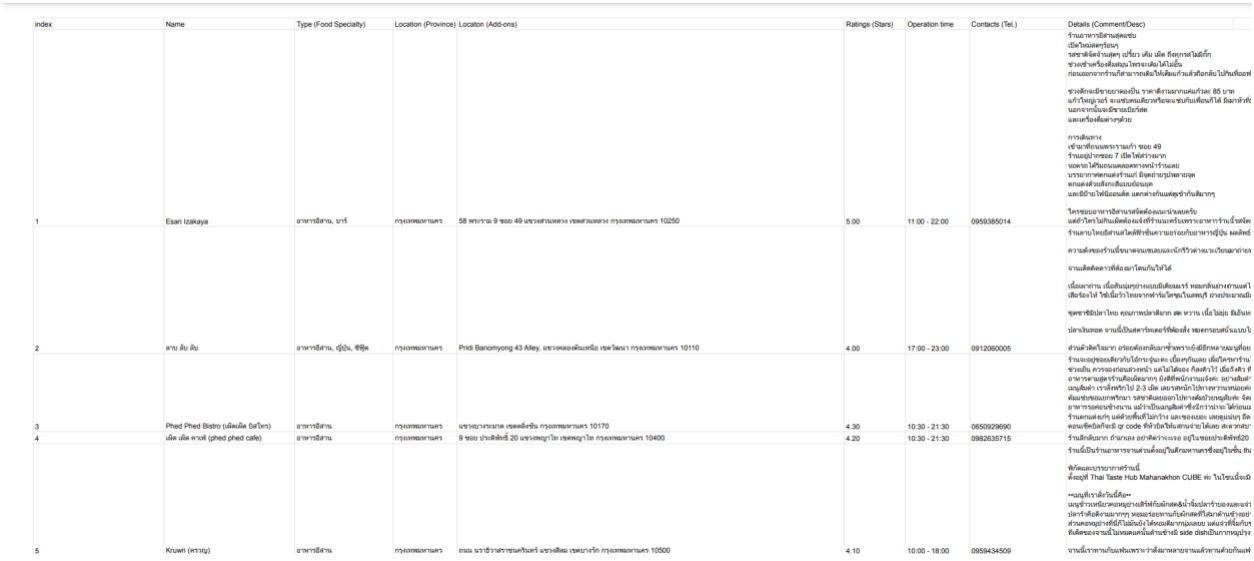
3.Mapping Platforms

-Google Maps

-Apple Maps

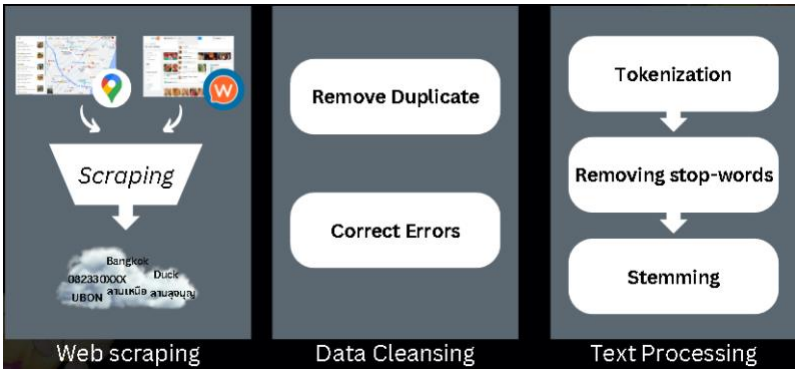
<u>Our data set for Esan restaurants search</u>

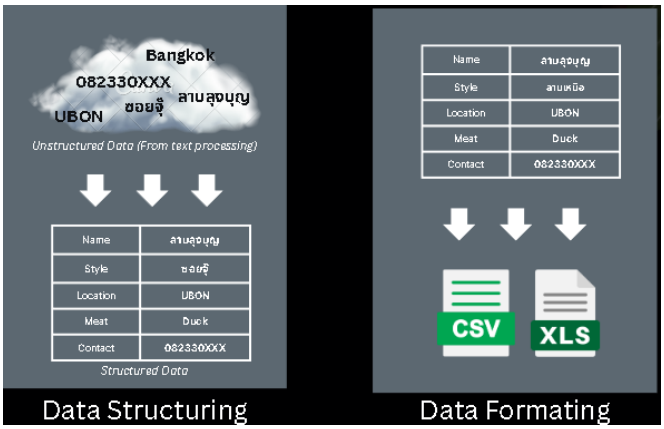
Mainly we have gathered from Wongnai and Google Maps and we put it into Microsoft Excel.



index	Name	Type (Food Specialty)	Location (Province)	Locaton (Add-ons)	Ratings (Stars)	Operation time	Contacts (Tel.)	Details (Comment/Desc)
1	Esan Izakaya	อาหารอีสาน, บาร์	กรุงเทพมหานคร	58 พระรามที่ 9 ซอย 49 แขวงสวนหลวง เขตสวนหลวง กรุงเทพมหานคร 10250	5.00	11:00 - 22:00	0959385014	
2	ลาบ ดับ ดับ	อาหารอีสาน, ญี่ปุ่น, ซีฟู้ด	กรุงเทพมหานคร	Pridi Banomyong 43 Alley, แขวงคลองตันเหนือ เขตวัฒนา กรุงเทพมหานคร 10110	4.00	17:00 - 23:00	0912060005	
3	Phed Phed Bistro (เผ็ดเผ็ด บิสโทร)	อาหารอีสาน	กรุงเทพมหานคร	แขวงบางรักพัฒนา เขตตลิ่งชัน กรุงเทพมหานคร 10170	4.30	10:30 - 21:30	0650929690	
4	เผ็ด เผ็ด คาเฟ่ (phed phed cafe)	อาหารอีสาน	กรุงเทพมหานคร	9 ซอย ประดิษฐ์มนูธรรม 20 แขวงคลองจั่น เขตบางกะปิ กรุงเทพมหานคร 10400	4.20	10:30 - 21:30	0982635715	
5	Kruwn (ครวญ)	อาหารอีสาน	กรุงเทพมหานคร	ถนน พระรามที่ 4 แขวงสีลม เขตบางรัก กรุงเทพมหานคร 10500	4.10	10:00 - 18:00	0959434509	

How to Pre-process our Data

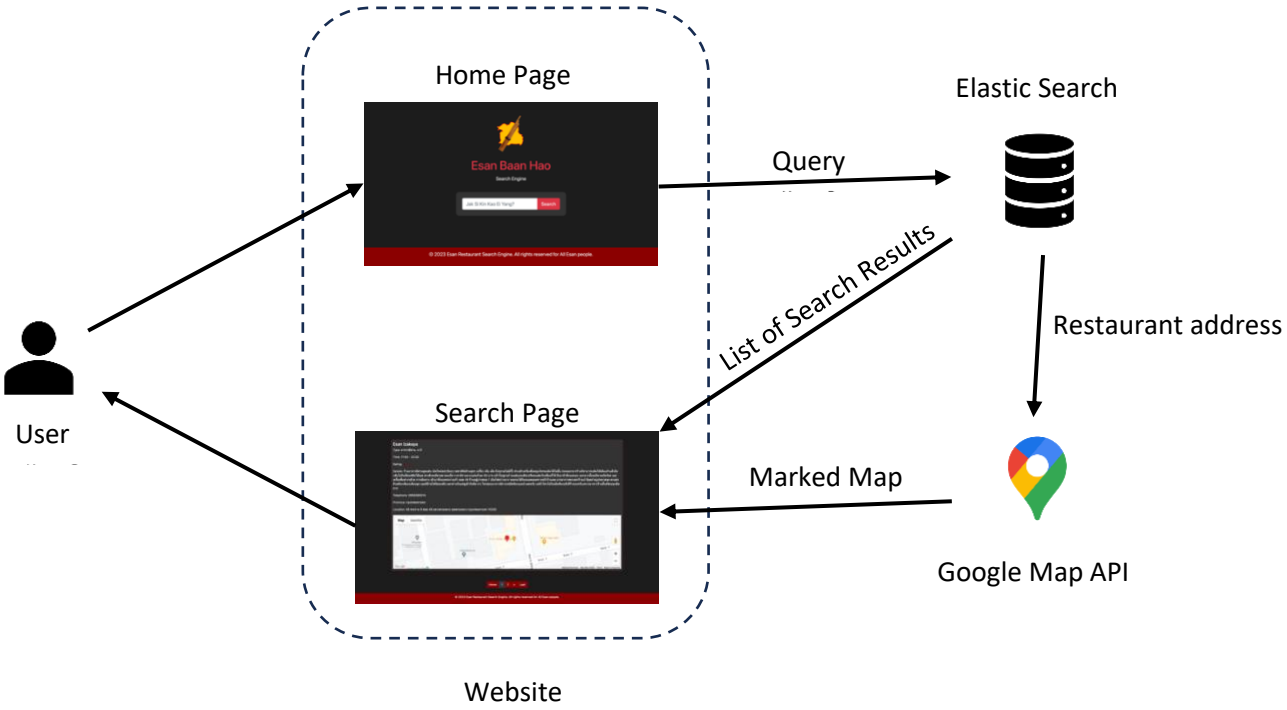




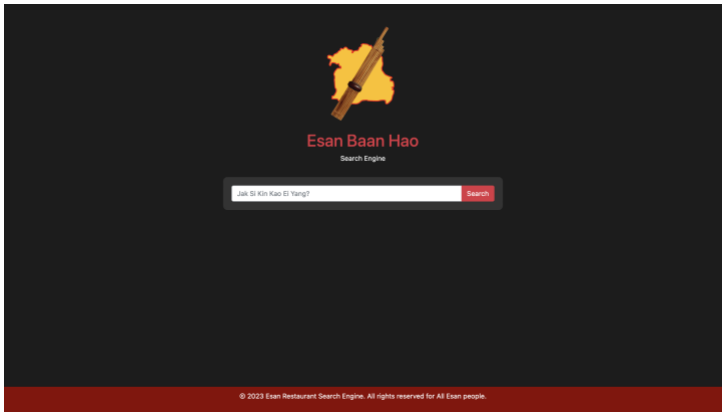
ii. Tools and software

- Elasticsearch
  - It is used to create the index and implement the IR system.
- Kibana
  - It is used for interface and interacts with Elasticsearch.
- FLASK
  - It used to create and host the search engine website.
- Google Map API
  - It is used to display the restaurant location on the Google Map
- Bootstrap
  - It is used for styling the website.

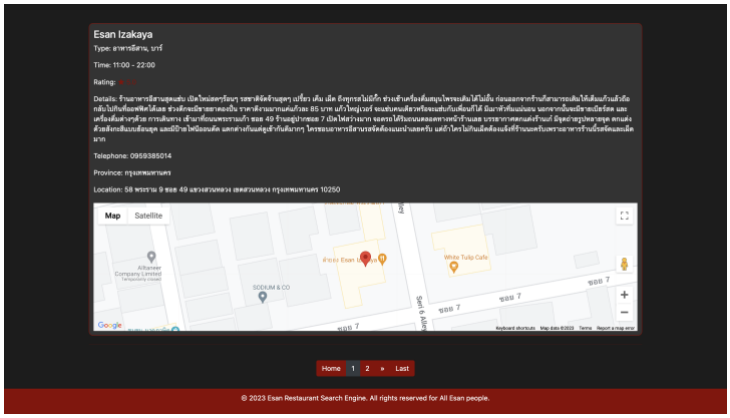
iii. System diagram



iv. Snapshots of the system



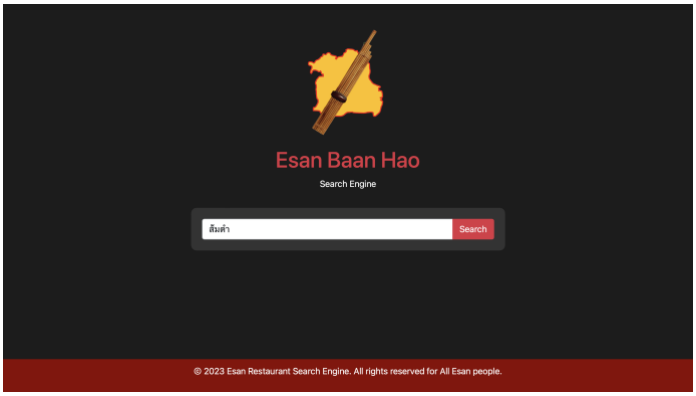
Screenshot of Home Page



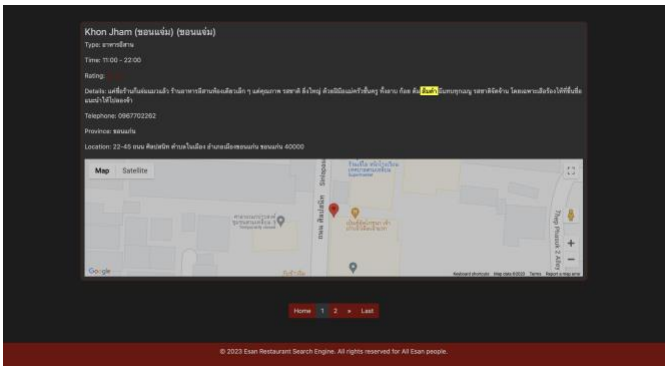
Screenshot of Search Page

v. Example step-by-step search sessions that highlight the following functionality:

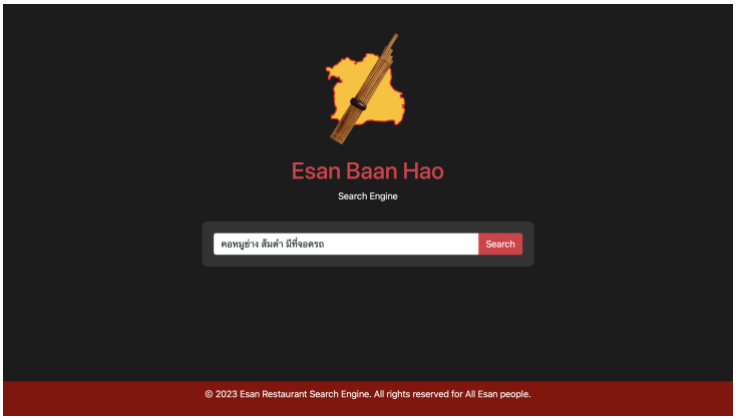
1. One word query
    - Query: “ส้มตำ”



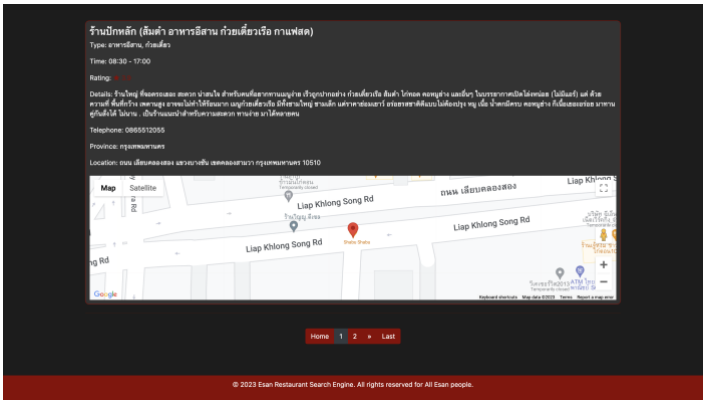
    - Search result



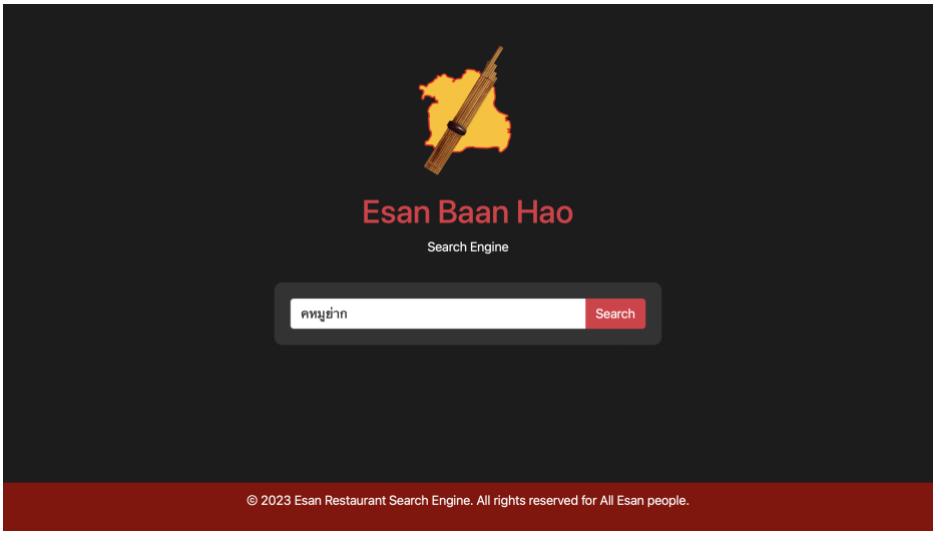
2. Multiple word query
    - Query: “คอหมูย่าง ส้มตำ มีที่จอดรถ”



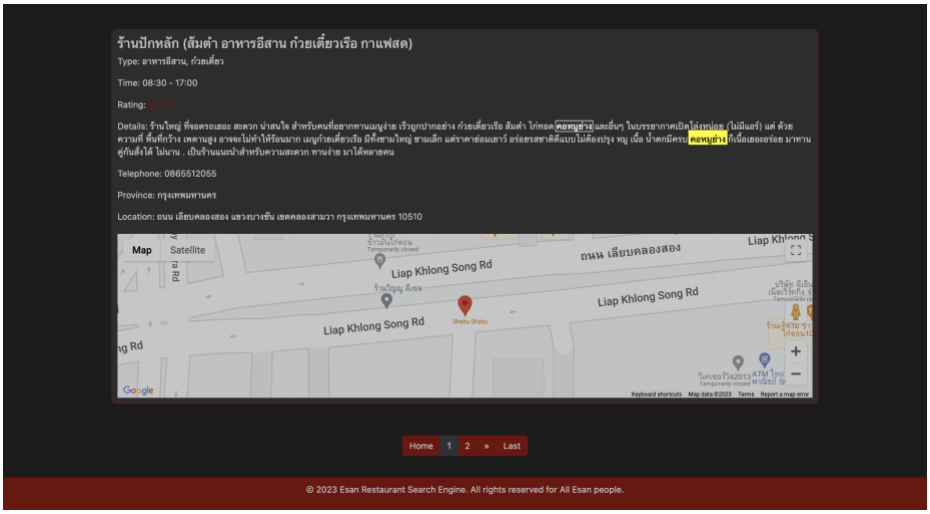
    - Search result



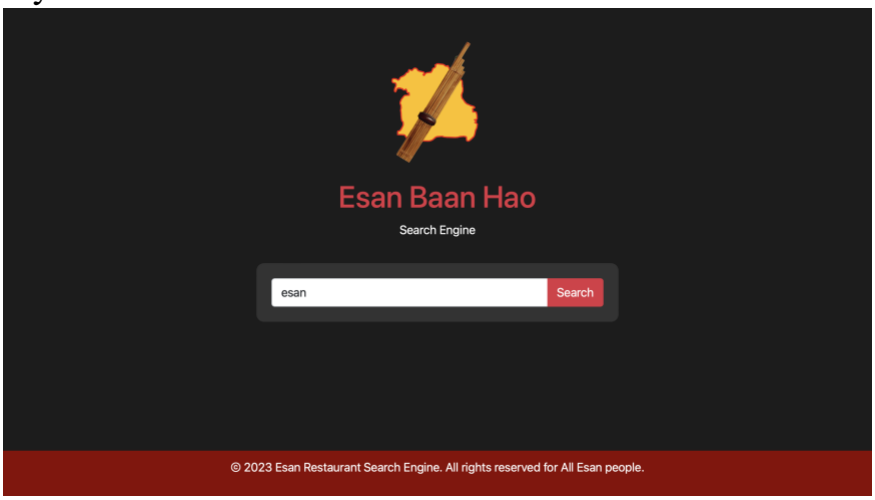
3. Partial match
    - Query: “คหมูย่าก”



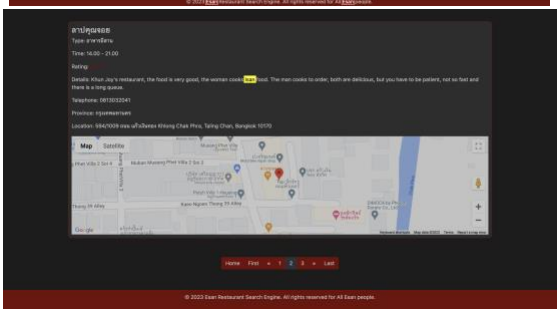
    - Search result

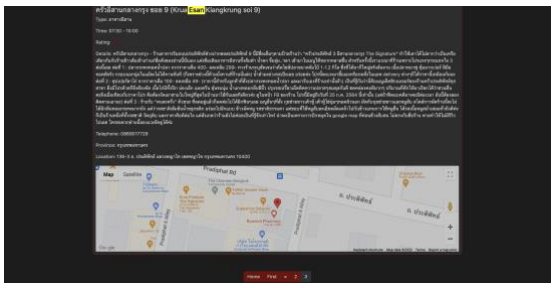


4. Ranking
    - Query: “esan”



    - Search result





3.Discussion

Limitations of your system

The size of the dataset is one of our limitations. In the current system, there are only 100 restaurants in our Esan restaurant search engine. And there is no the filtering system for adjusting the search characteristic.

Technical difficulties, challenges, and lessons learned.

Data sets related to the topics.

- We decided to choose “larb” as our main topic for the search system. Due to limited data to design the system. We’ve decided to change our topic to eastern food or “Esan” food instead.

Noise in data set

- Watermarked information (link attached Text/Paragraph)
- Emoji-Special characters in Texts/Paragraph

This can reduce the searching effectiveness

Word Segmentation Issues

- TF_IDF based algorithm.
- Continuous Script (No space between words).

Our search engine designed to support both English and Thai languages. However, TF_IDF based algorithm can’t support continuous script language like Thai. So, we tackle this through tokenization.

Opportunities for future improvements

Enhance Search Options

- Add more search option like food type, location, and opening hour

Expand dataset

- Import more restaurant into data set with more information

4.Conclusion

In conclusion, the Esan restaurant Search engine prototype is designed for people who want to find a good Esan restaurant tailored to their needs. The user can easily see the reviews and the location through Google Maps API.

The system can handle one-word queries, multiple-word queries, partial matches, and ranking functionalities to demonstrate the capability of the IR system.

However, there are limitations and challenges with our search engine, like the limited size of the dataset or searching with continuous script, which we have overcome by tokenization technique.

Despite these challenges, the opportunity for future improvement is awiated us, with the addition of Search Options and the expansive dataset.