

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:

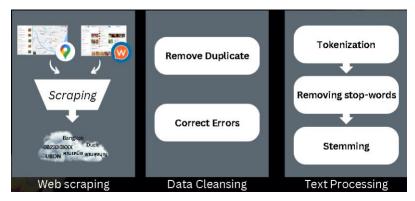
- 1.Restaurant Review platforms and websites
 - -Wongnai
 - -Michelin guide
- 2. Social Media Platforms
 - -Instagram
 - -Facebook
- 3. Mapping Platforms
 - -Google Maps
 - -Apple Maps

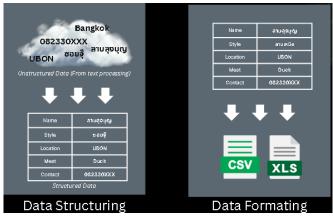
Our data set for Esan restaurants search

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

dex	Name	Type (Food Specialty)		Location (Add-ons)	Ratings (Stars)	Operation time	Contacts (Tel.)	Details (Comment/Desc)
								ร้างอาการที่สายสมเหน้า เป็นก็จะได้เหน้าสู่สาย เราะ เดียร์การสาย เราะ เดิน เด็น โดย การกำเนิดโกก เกราะ เด็นร้างสุดๆ เราะ เด็น เด็น โดย การกำเนิดโกก เกราะเราะทำเราะ เด็น เราะ เด็น โกกันแล้ว เด็น โกกัน เกราะ เด็น โกกัน เกราะ เด็น โกกัน เกราะ เกราะ เกราะ เด็น โกกัน เกราะ เกราะ เด็น โกกัน เกราะ เก
	Esan Izakaya	อาหารอิสาน. มาร์	024009000000	58 พระกาม 9 ช่อย 49 แนวงส่วนหลวง เขตส่วนหลวง กรุงเทพมหานคร 10250	5.00	11:00 - 22:00	0959385014	ใครขอบอาหารอีสานรสจัดต้องแนะนำเลยครับ แต่อำใครไม่กินเด็ดต้องแจ้งที่ร้านนะครับเพราะอาหารร้านรั
	Cour (Caraya			90 HILL & 200 -5 85 HILL HOLD HOLD HOLD HOLD HOLD HOLD HOLD HO	0.00	11.00 - 22.00	000000014	ร้านลาบไทยอีสานสไดล์ฟัวขึ่นความอร่อยกับอาหารญี่ปุ่น ค
								ความดังของร้านนี้ขนาดจนเซเลยและนักวิวิวต่างแวะเวียนม
								จานเด็ดติดดาวที่ต้องมาโดนกันให้ได้
								เนื้อเคาก่าน เนื้อสับเล่มๆอ่างแบบมีเดียลมาร์ พอมกลิ่นอ่างถ่า เสือร้องไห้ ใช่เนื้อวัวไทยจากทำรัมโดชุมในลพบุรี อ่างประเ
								ชุดชาชีมีปลาไทย คุณภาพปลาดีมาก สด หวาน เนื้อไม่ยุ่ย รั ปลาสับทอด บานนี้เป็นสดาร์ทเตอร์ที่ต้องตั้ง พอดกรอบสนับ
	ความ คับ คับ	อาหารอิสาน, ญี่บัน, ซีฟีด	กระเทพมหานคร	Pridi Banomyong 43 Alley, เหรวงคลองลับเหนือ เชลาัฒนา กวรเรพณหานคร 10110	4.00	17:00 - 23:00	0912060005	ส่วนตัวสิดใจมาก อร่อยต้องกลับมาข้าเพราะยังมีอีกหลายเม
			re georgeal M TARPS		1.00	17.50 - 23.00	W112/00AA	ร้านจะอยู่ขอยเดียวกันได้การจุ่นเคย เนื้องๆกันแลย เมื่อใควห ข่ายเป็น ควาจจะกัดมหัวหน้า แล้นไม่โด้ยอง ก็กลักวิรี ใน้้อนี้ อาหารหายสุดรู้จะก็เพื่อเดียมกาย เพื่อที่ทำนักวาลแล้วเล่ะ อย่า แม่ผู้แต่น เวาที่สาทักไม่ 23 เมิด และเกตเก็บไม่การการกาย เพียนข่ายขอยแกรที่ขอก ราสหายิดเลอกไปทางต้องก็วาลหุดให้เอา อาหารจะต่อเหรือรามาน แม้ว่าเป็นเกลูติดเท่ารี่จะก็วารกายกลัง ร้านเลยแล้งครั้ง และเกิดเลย แล้วเล็บไม่และเลยได้เรื่อง
	Phed Phed Bistro (เด็ดเด็ด ปีสโทว)	อาหารอิสาน	บรัสเทษสหมภาพร	แชวงบางระมาด เขตคลั้งขัน กรุงเทพมหานคร 10170	4.30	10:30 - 21:30	0650929690	ดอนเชิดปิลก็จะมี qr code ที่หัวปิลให้แสกนจ่ายได้เลย สะด
	เห็ด เพิด คาเพี (phed phed cafe)	อาหารอิสาน	กรุงเทพมหานคร	9 ชอย ประดิพัทธิ์ 20 แขวงพญาไท เขตพญาไท กรุงเทพมหานคร 10400	4.20	10:30 - 21:30	0982635715	ร้านศึกดับมาก ถ้ามาเอง อย่าคิดว่าจะเจอ อยู่ในชอยประศิพั
								ร้างเน็บโปรากะทางางาหล่ายตั้งอยู่ในศึกษาการทางรี้จะยุใน ที่ก็ผมแรงระบากจำรับเลื่ ด้อยผู้ที่ไรกับ Taste Hub Mahanakhon CUBE ค่ะ ในโขร "สมเด็บราดิรักน์ศึกษา แม่ข้าวหนีราคายหมูข่างสักร์กับลักสตครั้นใช้แอโกรทำของ มากว่าที่ดีคริการการทุก ของสาร์ตการกับลอก แต่นว่าที่สามารถเลือน สามารถเลือนกรุกการทุก ของสาร์ตการกับลอก แต่นว่าที่สามารถเลือน แต่นว่าที่สามารถเลือน แต่นว่าที่สามารถเลือน แต่นว่าที่สามารถเลือน แต่นว่าที่สามารถเลือน แต่นว่าที่สามารถเลือน เลือนใหม่ในการการการการการการการการการการการการการก

How to Pre-process our Data

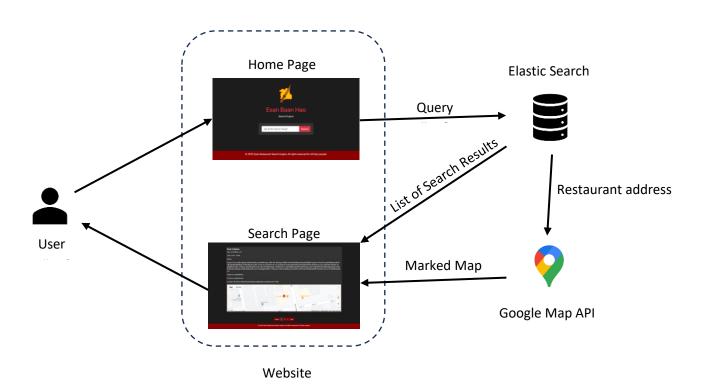




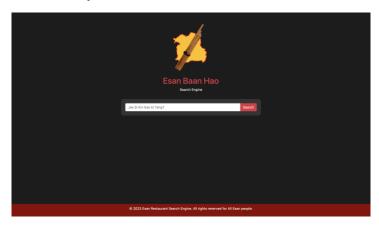
ii. Tools and software

- Elasticsearch
 - It is used to create the index and implement the IR system.
- Kibana
 - o It is used for interface and interacts with Elasticsearch.
- FLASK
 - o 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
 - o It is used for styling the website.

iii. System diagram



iv. Snapshots of the system



Screenshot of Home Page

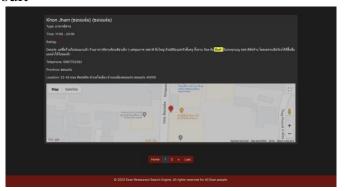


Screenshot of Search Page

- v. Example step-by-step search sessions that highlight the following functionality:
 - 1. One word query
 - Query: "ส้มตำ"



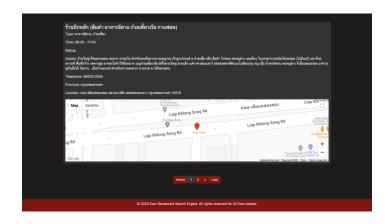
• Search result



- 2. Multiple word query
 - Query: "คอหมูย่าง ส้มตำ มีที่จอครถ"



• Search result

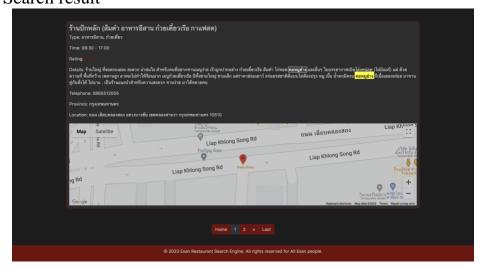


3. Partial match

• Query: "คหมูย่าก"



• Search result



4. Ranking

• Query: "esan"



• Search result



3.Discussion

<u>Limitations of your system</u>

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.