

# **SNACK HUNTER**

## **THE ULTIMATE TAIWANESE NIGHT MARKET FOOD FINDER**

**四234 - GROUP 9 - MIS**

112306079 彭宇承 112306092 許新翎 112306093 林秉毅  
112306018 陳詠旭 112701030 陳允中

# 1131 Data Structure – Final Project Report

四 234 – Group 9 – MIS

112306079 彭宇承 112306092 許新翊 112306093 林秉毅 112306018 陳詠旭 112701030 陳允中

## Final Project – Let's Beat Google!

### I. Introduction

#### A. Topic

**Snack Hunter: The Ultimate Taiwanese Night Market Food Finder**

#### B. Motivation

In this rapidly evolving technological era, people are increasingly relying on the convenience brought by advanced technologies. However, effectively leveraging these tools has become a critical challenge. Taiwan's night markets are globally renowned for their diverse and unique culinary delights. Yet, for international visitors, language barriers, religious dietary restrictions, or personal preferences often make it difficult to accurately identify the ingredients and reviews of the food in front of them.

To address this issue, we have developed a unique search technology that integrates language translation with artificial intelligence. Our system intelligently recognizes user input keywords and food descriptions, converting them into optimized keywords suitable for Google searches. By leveraging Google search results, combined with our extensive database of Taiwan night market food-related keywords, the system maps, re-scores, and reorders the results to enhance their accuracy. This technology empowers travelers from around the world to overcome language and cultural barriers, enabling seamless exploration of Taiwan's diverse night market cuisine. Not only does it allow visitors to enjoy local delicacies with greater confidence, but it also significantly enhances Taiwan's cultural visibility and accessibility on the global stage.

## C. Why can we beat Google? (Our Advantage)

### 1. Multilingual Processing

When searching for "Taiwan night market food" using foreign languages, the results are often irrelevant or entirely absent. To address this, we propose a "query preprocessing" framework that integrates a translation API and a classification model to process queries more effectively. In this approach, user queries are first translated into Chinese and then categorized into three key categories:

- County/City
- Night Market Name
- Food Name

For the **County/City** and **Night Market Name** categories, instead of directly using the translated terms, the system compares them against predefined lists of all Taiwanese cities and night markets using a similarity-matching model (e.g., Voyage Re-ranker). The top-ranked match is selected and used to replace the user-provided term with the closest valid option from the list. This ensures consistency and accuracy in identifying specific locations or night markets.

By leveraging this targeted translation and similarity-checking process, combined with the classification model, our framework enhances flexibility and precisely maps user intents. The refined and categorized keywords are then used to perform a search in Chinese, significantly improving search precision compared to Google's native capabilities for non-Chinese languages.

### 2. Natural Language to Keywords Conversion

Using CKIP Transformer technology, we preprocess user input to convert both natural language queries and keyword-based questions into a structured keyword format. This approach allows for flexibility in input styles, eliminating the need for users to follow Google's rigid keyword-based query requirements.

### 3. Relevance Reordering with Localized Weighting

Our system analyzes a vast collection of Taiwan night market-related keywords, applying a weighted scoring mechanism to re-rank Google search results. Unlike Google's generic approach, which lacks customization for niche topics like Taiwan night market cuisine, our localized keyword database enables enhanced relevance. This ensures that search results are significantly more accurate and tailored to the user's intent.

## II. Search Tricks

### A. Keywords (Our system will classify the user's input into one of the three categories using a classification model after the input is provided.)

**County/City :** 臺北、高雄、臺中、...

**Night Market Name :** 士林夜市、逢甲夜市、饒河夜市、...

**Food Name :** 鹽酥雞、臭豆腐、大腸包小腸、...

#### Keyword Weighting:

**County/City :** 3.0

**Night Market Name :** 5.0

**Food Name :** 15.0

### B. Score Formulation

#### 1. Keyword Matching Scoring

When the user's query keywords match the content on a website, a base score is assigned based on the frequency of the keywords appearing on the site. This score is adjusted according to the degree of keyword match.

Unlike assigning individual weights to single keywords, our keyword matching scoring system assigns fixed weights based on keyword categories (e.g., county, night market name, etc.).

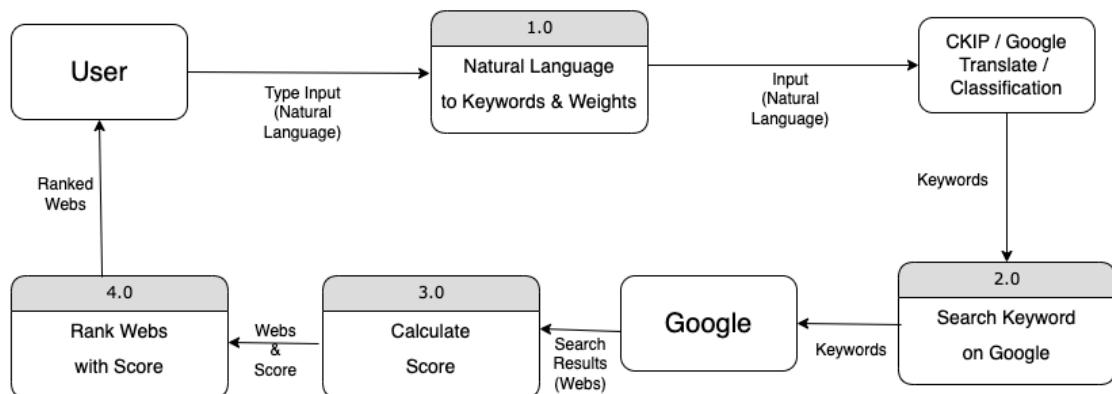
## 2. Weighted Calculation

$$\text{Total Score} = (\text{Number of County Matches} \times 3) + (\text{Number of Night Market Name Matches} \times 5) + (\text{Number of Food Name Matches} \times 15)$$

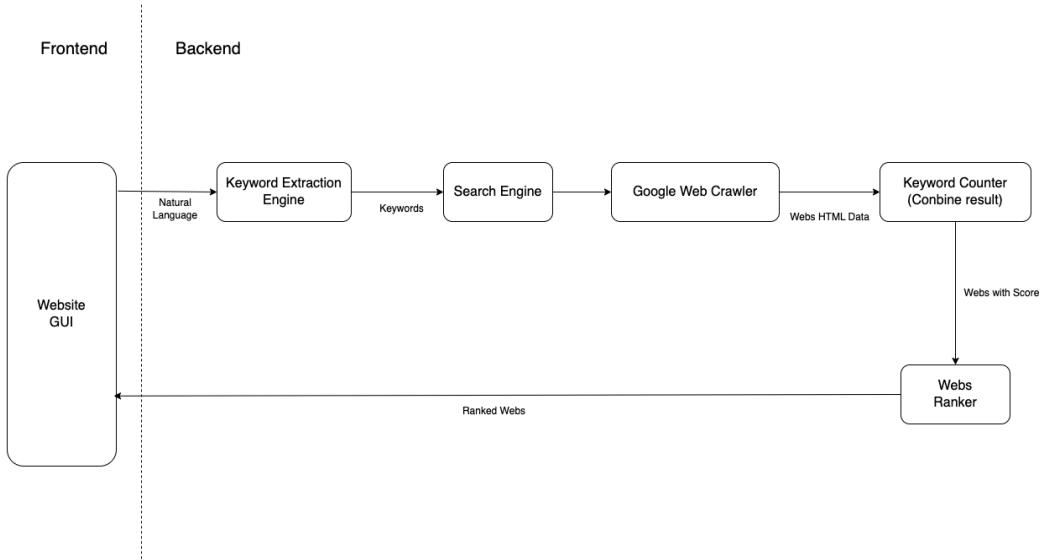
## III. System design

- Source Code: [https://github.com/RoyPeng126/DS\\_Project](https://github.com/RoyPeng126/DS_Project)
  - Setup environment description file, see README.md
- Slides:  
[https://github.com/RoyPeng126/DS\\_Project/blob/main/DSppt\\_Group9\\_final.pdf](https://github.com/RoyPeng126/DS_Project/blob/main/DSppt_Group9_final.pdf)

### A. Data flow diagram



## B. System structure diagram



## C. Deep Dive: Keyword Extraction Engine

- **Query Preprocessing framework**
  - integrates a **translation API**, a **CKIP model** and a **classification model**
- **Classification Model:** We precompile a comprehensive list of all "cities" and "night market names" in Taiwan. Using the Voyage ReRanker, the user's input keywords are compared with all entries in the list based on text similarity.
  - **If the highest similarity score is above 0.5**, the input is classified according to the corresponding category of the matched value, and the user's input is replaced with the matched entry in the list.
  - **If the highest similarity score is below 0.5**, it indicates that the input does not relate to any city, or night market name. In this case, the input is directly classified under the "food name" category.

A screenshot of a JSON response from a Voyage Reranker API. The response shows a query for "石林夜市" and a list of documents including "士林夜市", "六合夜市", "臺北市", "新北市", "桃園市", and "桃園觀光夜市". The reranked results show "士林夜市" with a score of 0.5703125 and "桃園觀光夜市" with a score of 0.314453125.

```
1 {
2   "query": "石林夜市",
3   "documents": ["士林夜市", "六合夜市", "臺北市", "新北市", "桃園市", "桃園觀光夜市"]
4 }
```

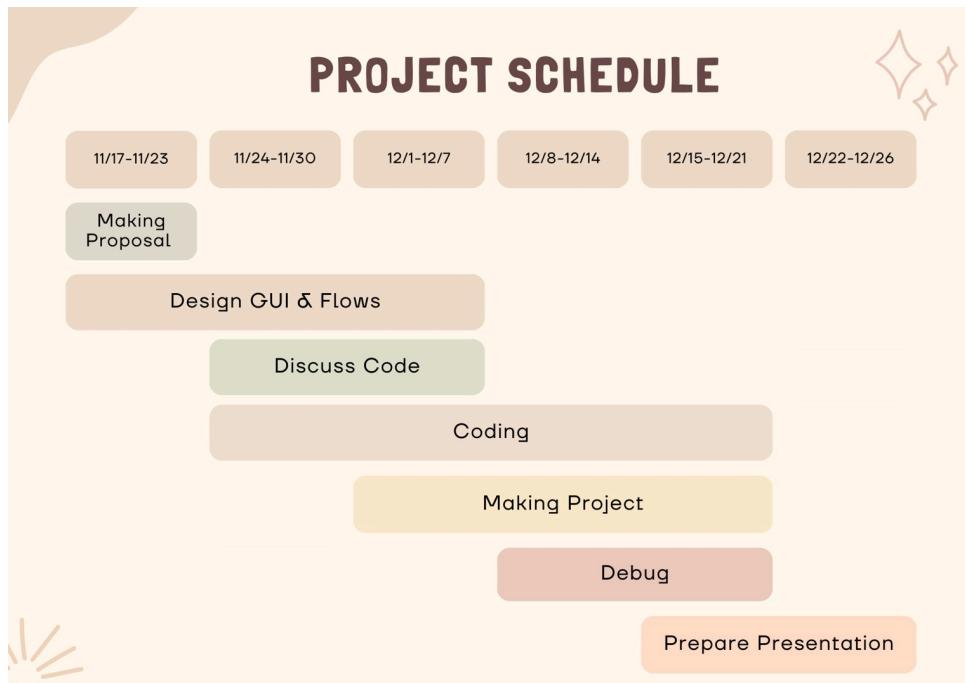
Body Cookies Headers (5) Test Results | ⏱

Pretty Raw Preview Visualize JSON ↻

```
1 {
2   "ranked_documents": [
3     {
4       "document": "士林夜市",
5       "score": 0.5703125
6     },
7     {
8       "document": "桃園觀光夜市",
9       "score": 0.314453125
10    }
]
```

Voyage Reranker Output Example Diagram

## IV. Schedule



## V. Challenges

### A. Translation Errors

Due to cultural differences, a single word may refer to different foods in different regions. For instance, "chips" in British English means "fries," while in American English, it refers to "potato chips." If a user's search query contains such terms, it may result in incorrect outputs.

### B. Precision of Word Usage

If users are unsure how to describe the snack they wish to explore and use vague terms (e.g., color, shape, flavor), the search engine may output results that match the characteristics but may not align with the exact snack the user is looking for.

### C. Search Feedback Speed

To enhance the search accuracy for this specialized topic (Taiwan night market cuisine), we integrated CKIP Transformers for tokenization. However, combining these processes with the Google API may result in longer execution times for a single search, potentially affecting the user experience.

## VI. System Demo

### 1. User Interface (Web)

The image is a collage of three photographs. On the left and right are night market scenes with various signs and people. In the center is a screenshot of the NMSL (Night Market Search List) web application. The app has a search bar with Korean text "저는 류하 야시장에서 파파야 밀크를 사고 싶어요." (I want to buy papaya milk at Ryehwa night market). Below it is a blue "Search" button. The main content area shows a query "Query: 六合觀光夜市 繁華夜市 木瓜 牛奶". It displays two search results. The first result is a box with the title "高雄六合觀光夜市攻略 : Top12六合夜市美食、必逛景點、伴手禮推薦 ..." and a snippet about木瓜牛奶 (Watermelon Milk). The second result is a box with the title "[2024高雄六合觀光夜市10間美食推薦] 美食雲集高雄必逛六合夜市 ..." and a snippet about木瓜牛奶 (Watermelon Milk). Both results have a "Score: 645" and "Score: 475" respectively. Red circles highlight the text "木瓜牛奶" in both snippets.

### 2. User Interface (iOS)



### 3. Related Keywords (Derive relative keywords by top-ranked websites)

We directly scrape all the content from the 'People also search for' section at the bottom of the 'Google Search Results Page' to serve as the information for this block.

