

Final Project Report

Wengxi Li (wengxili@umich.edu)

1. Project Code

All the project code and documentations can be found at this GitHub repository (https://github.com/imerlwx/SI507_Final_Project).

This repository includes all the data source cache (1 .json file), texts (2 .txt files) and source codes (9 .py files) of SI 507 final project. No need of API keys for the tour guide system because most of the data have been stored in cache. It will be necessary to apply a Yelp Fusion API for the restaurant recommendation system. If you do not apply a new key, the system will use the default one. Python packages this program uses are requests, json, itertools, sys, collections, and webbrowser. Most are built-in packages. To run the program, just execute the Interface.py and follow the interacting instructions shown.

2. Data sources

The first data source is from the Distance Matrix. Sample data and documentation can be found at this URL (<https://distancematrix.ai/product>). The formats of this data source will be JSON. The data can be accessed by using API keys, which can be applied on the authentication website for developers (<https://distancematrix.ai/contact>).

We will use this API to get the distance between two arbitrary locations, so it is proper to store the data into caching file. The file has 300 records. The most important field is the 'distance', 'origin_address' and 'destination_address'. The evidence of caching is shown in one snapshot of the cache file as Figure 1.

The second data source is Yelp Fusion. Sample data and documentation can be found at this URL (https://www.yelp.com/developers/documentation/v3/business_search). The formats of this data source will be JSON. The data can be accessed by using API keys, which can be applied on the authentication website for developers

(<https://www.yelp.com/developers/documentation/v3/authentication>).

Because we only access this API for a small number (always 10) each time, there is no need to use cache. For each record, the attributes are shown in Figure 1. For our specific purpose, the ‘name’, ‘rating’ and ‘url’ are the most important fields. Different with the figure shown, we will use the data from Ann Arbor.

[illegible]

Figure 1 Part of the data source cache file

```
{
  'alias': 'golden-boy-pizza-san-francisco',
  'categories': [{'alias': 'pizza', 'title': 'Pizza'},
                 {'alias': 'italian', 'title': 'Italian'}],
  'coordinates': ({'latitude': 37.7997956, 'longitude': -122.4080729},
                 {'display_phone': '(415) 982-9738',
                  'distance': 4812.020663297437,
                  'id': 'PTfextXS47ZVRCDzRfEWGw',
                  'image_url': 'https://s3-media1.fl.yelpcdn.com/bphoto/SfosPh2kVikKALLVtL8g/o.jpg',
                  'is_closed': False,
                  'location': {'address1': '542 Green St',
                              'address2': '',
                              'address3': '',
                              'city': 'San Francisco',
                              'country': 'US',
                              'display_address': ['542 Green St', 'San Francisco, CA 94133'],
                              'state': 'CA',
                              'zip_code': '94133'},
                  'name': 'Golden Boy Pizza',
                  'phone': '+14159829738',
                  'price': '$',
                  'rating': 4.5,
                  'review_count': 3907,
                  'transactions': ['delivery'],
                  'url': 'https://www.yelp.com/biz/golden-boy-pizza-san-francisco?adjust_creative=06u090-ny2pC05uU0BL0&utm_campaign=yelp_api_v3&utm_medium=api_v3_business_search&utm_source=06u090-ny2pC05uU0BL0'})
}
```

Figure 2 A sample of data from the Yelp Fusion API

3. Data Structure

The basic data structure will be a graph. Each node of the graph represents a destination that user want to visit. Each edge represents the distance of the two locations. The graph will be depicted using the Graph class, which defines some attributes and methods. The attributes include edge list, edge dictionary and edge matrix, as shown in Figure 3. The methods are used to get these three attributes. This structure will be stored in the cache file graph.json. To read this file, the ReadFile.py includes some method. One is to read all the file contents, one for reading the corresponding locations, one for getting the size of files, the last one for reading contents line by line. Use the first function is enough.

| graph.edgeList | graph.edgeMatrix |
|--|---|
| [(0, 2, 871), (0, 4, 1243), (0, 6, 1539), (2, 4, 362), (2, 6, 1084), (4, 6, 699)] | [[0, 871, 1243, 1539], [871, 0, 362, 1084], [1243, 362, 0, 699], [1539, 1084, 699, 0]] |
| graph.edgeDict | |
| defaultdict(list, {0: [2, 4, 6], 2: [0, 4, 6], 4: [0, 2, 6], 6: [0, 2, 4]}) | |

Figure 3 A schematic of the graph structure

4. Interaction and Presentation Plans

The interaction will be command line prompts. There will be three choices for user: tour guide system, local restaurant recommendation system, or exit the whole program.

In the tour guide system, user will choose the destinations from the University of Michigan's buildings with brief introduction. There are totally 25 destinations but to get results instantaneously it is better to only choose 10 or less destinations. After user choose the point to start, the shortest path and distance will be output. User will have an option to choose whether showing the graph structure or exit this system.

In the local restaurant recommendation system, user could input the sorting requirement, number of results to show, and the food category to search. The results will show the restaurants list. User could choose to see the webpages about each restaurant in the browser or just input other food category to start a new search or exit this system.

5. Demo Link

The demo video with narration descriptions can be found at this link:

<https://umich.zoom.us/rec/share/8Kft0TorWQBBvmilaFNQ-Fnj0EIFDLwnAs0payXnIfEFHUDaVtsMI0M3-uh09NL.QY2hOSCBsv5NMVWb> .

It shows how a user would interact with the program to perform different functions.