

DSCI558 Project Proposal

Project name: LA travels

Project members: Bin Zhang, Hanzhi Zhang

Project Domain & Goals:

We plan to build a geographic knowledge graph. The knowledge graph will have information about the place of interest (POI). We will focus on the POI related to the hotel, restaurants, attractions, and shopping stores. The attributes for each item are shown below:

Hotel: name, rate, location, average price, environment, room type, and service.

Restaurant: name, location, menu, average price, environment, taste, and service.

Attraction: name, location, attraction types (nature or man-made), rate, picture, features (entertainment, high-rise building, or art building).

Shopping store: item type, average price, the flow of people, ranking rate.

We will not only crawl the basic information but also process NLP on reviews. Therefore, there are also many attributes that might be discovered during information extraction. For example, if there are many comments saying, they received discrimination from waiters. We might add a new attribute, racism (NO/Possible/Yes).

If the knowledge graph is established, an intelligent question answering (QA) will be used to help people building a wonderful LA travel that satisfies the requirement from users. The current travel recommendation models based on the knowledge graph are always focusing on the macro perspective, which means the model will focus on which city to go. However, our model will give the recommended places from multi requirements from users and design a travel plan for him. The main reason is we added a distance prediction in our project.

Datasets:

This knowledge graph will contain data from Yelp, TripAdvisor, Booking, and Airbnb including basic information and reviews. We will also use Google API including Place API, and Distance Matrix API.

Yelp: https://www.yelp.com/search?find_desc=Restaurants&find_loc=Los+Angeles%2C+CA&ns=1

TripAdvisor: https://www.tripadvisor.com/Attractions-g32655-Activities-Los_Angeles_California.html

The above links are two example of seed URLs that we will use. There is no existing ontology for board games, so we will design a custom ontology.

Technical Challenge:

A basic problem of geographic knowledge graph is that compare with the typical geospatial map, it cannot represent the location relationship intuitively. For current most of the travel recommendation model, they use a predict, :nearby, to show this place is in this city. This can result in a problem that for attraction A and attraction B, even if they are both in the same city and recommended from the model, the distance between them might be very large. In order to represent the distance value. We create a distance entity and connect two different POI.

The other problem is entity resolution. If we find a review in a restaurant is that the shrimp is too spicy, we might need to find the shrimp plate in the menu and added a value spicy into the taste attributes of this food. We will LSTM or any possible method to solve it.