



Visualization of Yelp's Academic Dataset

Sun-Yi Lin(sl3833) & Hao-Hsiang Chuang(hc2751)

I. Introduction

Yelp is a famous recommendation service base on various businesses, it helps people to find great local businesses including restaurant, dentists, hair stylists, mechanics, etc. It was founded in 2004 and now it has 135 million monthly visitors and 71 million reviews. With the rapid growth of their dataset, Yelp now provides various online open source for people to develop innovative usage and design of them.

II. Motivation

When we first discuss what to do in the final project, visualization quickly become our consensus. Graph is a powerful tool that could make complicated records easy to read and understand. Yelp's Academic Dataset is the best choice to implement visualization. It provides all the data and reviews of the 250 closest businesses for 30 universities for students and academics to explore and research, and we can do so much manipulations on it. We want to provide a interesting virtualization front-end application for the Yelp's Academic Dataset.

III. Tools & Library Description

a. Python

We use Python to parse the raw dataset to our desired json format.

b. JavaScript with D3 library

We use JavaScript with D3 (Data-Driven Document) library to build the visualization. D3.js is a JavaScript library for manipulating documents based on data. D3 helps us bring data to life using HTML, SVG and CSS. D3's emphasis on web standards gives us the full capabilities of modern browsers without tying ourselves to a proprietary framework, combining powerful visualization components and a data-driven approach to DOM manipulation.

c. jQuery

We use the jQuery library to make our HTML document easier to be handled with traversal, manipulation and animation.

d. Bootstrap

The basic framework of our front-end design.

IV. System Overview

We consider the following two graph models as our visualization target.

a. Sunburst Model



In sunburst model, we want to design an interactive graph that let the user to search for desired restaurants by the desired school campus.

b. Map Model



In map model, we want to design an interactive graph that let the user to search for desired restaurants by the desired location.

V. Design Concept

a. back-end part (Data Processing using Python):

Input :Yelp_academic_dataset.json, about 470 thousands objects

EX. Business Objects

```
{
  'type': 'business',
  'business_id': (a unique identifier for this business),
  'name': (the full business name),
  'neighborhoods': (a list of neighborhood names, might be empty),
  'full_address': (localized address),
  'city': (city),
  'state': (state),
  'latitude': (latitude),
  'longitude': (longitude),
  'stars': (star rating, rounded to half-stars),
  'review_count': (review count),
  'photo_url': (photo url),
  'categories': [(localized category names)]
  'open': (is the business still open for business?),
  'schools': (nearby universities),
  'url': (yelp url)
}
```

Output: json object [(academic_location, food_style, business_id, photo_url, review_count, restaurant_name, url, avg_stars)]

- b. front-end part (API using JavaScript with D3 Library) :

A. Sunburst Model

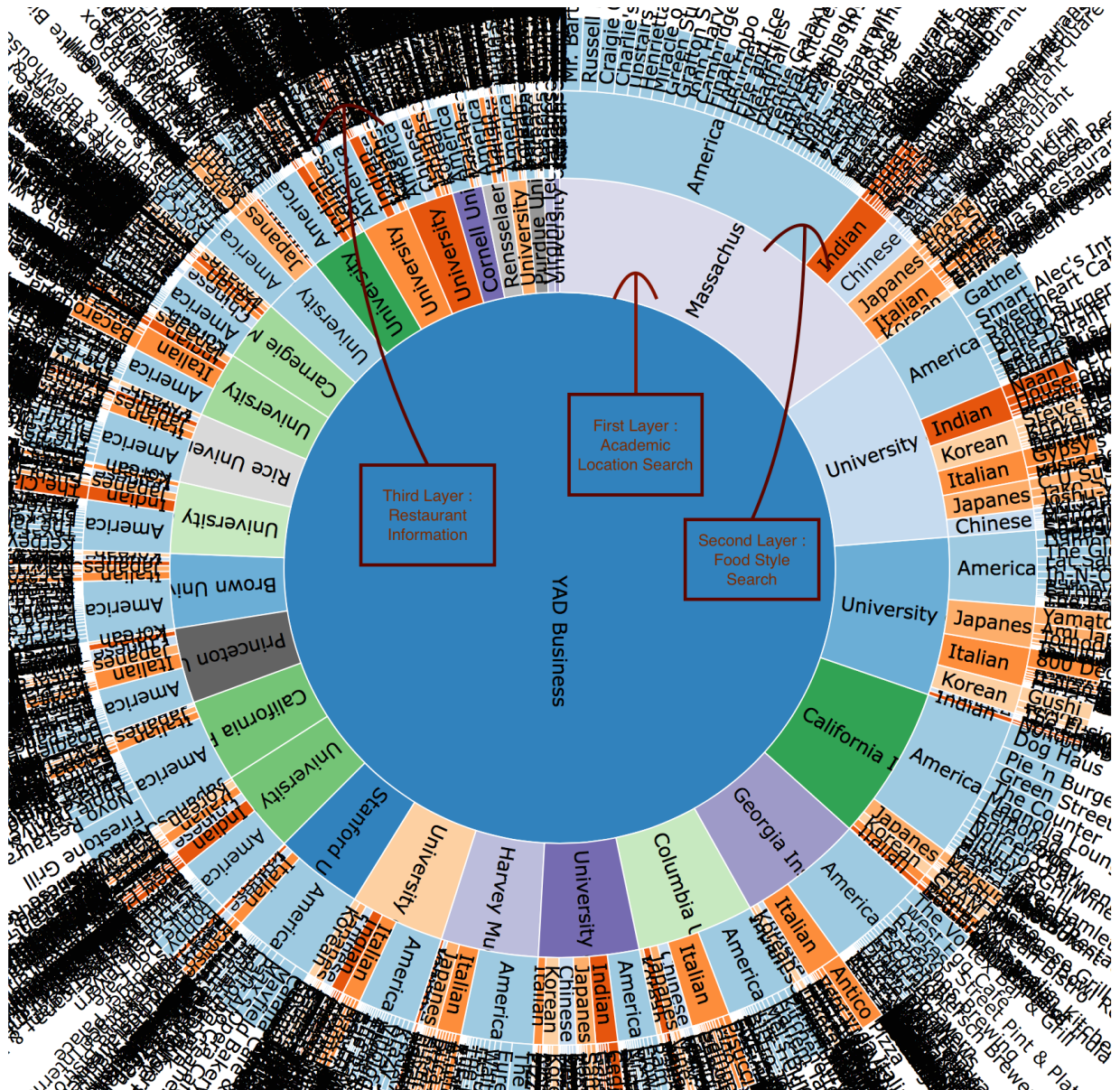
- a. First layer -- Academic Location Search
- c. Second layer -- Food Style Search
- d. Third layer -- Restaurant Information

B. Map Model

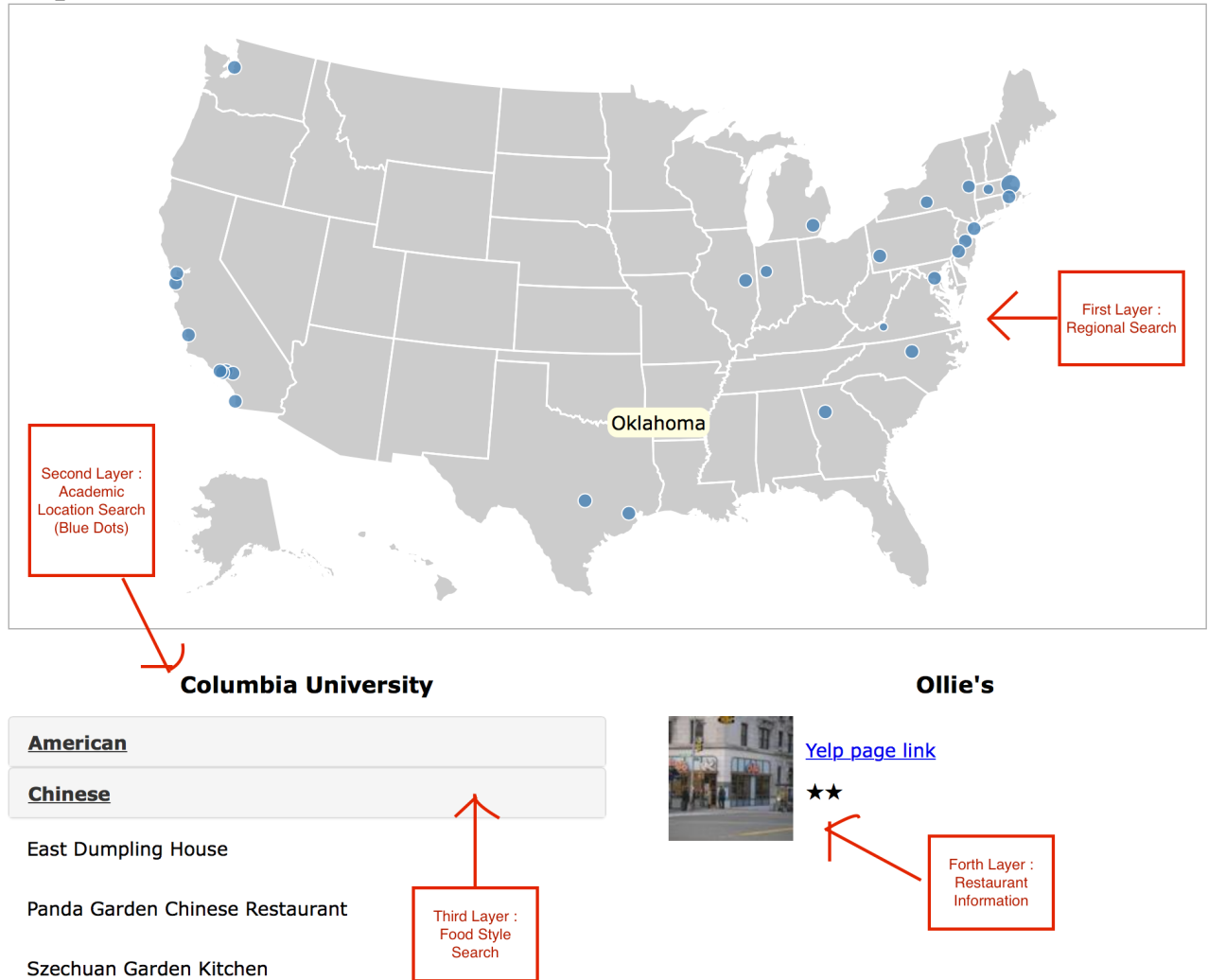
- a. First layer -- Regional Search
- b. Second layer -- Academic Location Search
- c. Third layer -- Food Style Search
- d. Forth layer -- Restaurant Information

VI. Results

a. Sunburst Model



b. Map Model



VII. References

Yelp Academic Dataset https://www.yelp.com/academic_dataset

D3 Library and Example <http://d3js.org/>

jQuery <https://jquery.com/>

Bootstrap <http://getbootstrap.com/>