Capstone Two Project Overview

Search engines are everywhere such as in e-commerce, social media, education, and general sites and they are frequently used daily by almost everyone around the world. Although restaurant search engines aren't unique and exist already in Yelp and other business review sites; I always wondered how search engines are built and processed. This project is my take on populating accurate restaurant listings using multiple datasets through natural language processing which I aim to provide specific results based on user's search input whether it's general search term (i.e. 'I want Chinese food') or more specific (i.e. 'guacamole with picadillo).

Hypothesis

Can using Allmenus' menu description and Yelp's Reviews/Tips text values be able to identify a coherent list of restaurants based on user query?

Criteria for success

Success for this project would be the training of multiple doc2vec models and able to list coherent restaurant lists based on user query. For instance, when the query is "dim sum" (Chinese/Taiwanese cuisine), Doc2vec models will be able to list all Asian, more specifically Chinese related restaurants.

Scope of solution space

Yelp's Review and Tips data along with Allmenus' menu description/title data will be analyzed.

Constraints

- Yelp's dataset is only limited to 10 states with a limited list of cities. Allmenus' dataset does not contain all restaurants' menus listed in Yelp's dataset.

Stakeholders

Management at Lexus of Mishawaka and plant production management at the Lexus Corporate level.

Data Source(s)

In this project, I'll be using multiple datasets related to restaurants, food/ingredients, and menu items. Yelp's dataset is used for getting businesses, reviews, and tips data. Allmenus' dataset (web scraped) which contains restaurants' menu offerings (description) and its menu title.