CS 122 Project Proposal

Group Name: Tattarrattat

Group Members: Xibai Wang, Shucen Liu, Can Liu, Lily Zhifan Zhou

Project Theme: Natural Language Processing (inspired by Yelp Dataset Challenge)

Goal: We would like to facilitate the customers' restaurant searching process by presenting them with easy-to-read and all-encompassing information compiled from Yelp reviews. Essentially, we want to create a user interface where the users enter the name of a restaurant, and get back diagrams related to the advantages and disadvantages of restaurants and similar restaurants, what considerations contributed to these, and how they could make a more informed decision. This saves them the trouble of reading long and emotional reviews and expedites the decision-making process.

Data Source: Scrape restaurant data from Yelp Chicago

Tentative Approach:

1. Extract keywords from comments, find the positive and negative words/phrases used in the reviews to calculate the score of each review.

2. Identify the temporary and long-term problems of each restaurant. (After discussing with Dr. Chaudhary in Week 6, we decided not to do this any more.)

3. Group comments into different categories such as service, food quality, ambience, and price (location listed as a searching criterion).

4. From customers' reviews, measure the relative importance of each different consideration. (Instead, we chose to ask customers to enter the ranks for different criteria when searching.)

5. Generate an unbiased and/or personalized rating by adjusting the relative importance of each consideration. (We calculated the scores for each different category for the user's reference.)

6. Give recommendations of similar restaurants based on the customers' criteria ranking and visualize the data into a diagram as well as a table with additional information.

Timeline:

Week 4:

Proposal and presentation

Week 5:

Create database from the website

Week 6:

Extract positive and negative words from reviews; identify highlights of the restaurants

Week 7:

Group comments and develop algorithms for personalized rating

Week 8:

Test algorithms, start building interactive website

Week 9:

Build an interactive website