# Restaurant Ontology and Personalized Recommendation

CS270 Team Symmetra : Ahmed Attia (aattia), David Eng (dkeng), Leo Keselman (leonidk), Stephanie Tang (svtang)

## Introduction

Our problem task was to build an ontological reasoning system to recommend restaurants to a user, given their preferences, ranging from dietary restrictions, budget, location, and favored cuisines, in addition to user history.

## Data Set

We are using the ***Yelp Academic Dataset.*** It includes ~800k restaurants, with ~2.5M reviews. Additionally, there are ~500 category tags (Chinese, Fast Food, etc.) that are applied to each business.

These category tags form the basis of our ontology, as we learn how various categories are organized and relate to each other. Along with other properties (city, rating), constitute our data modeling's understanding of the world.

## Multi-Class Classification

We built a system to understand how powerful different ontological categories of features are to modeling user preference. For each user's list of reviews, we held out one review and tested whether we could predict the number of stars the user gave to this review, given the remaining reviews. We filtered for users with at least 50 reviews. Random guessing missed by an average of 1.14 stars

### *Using Business Categories*

Using the provided business category information, k-nearest neighbors missed by an average of 1.06 stars.

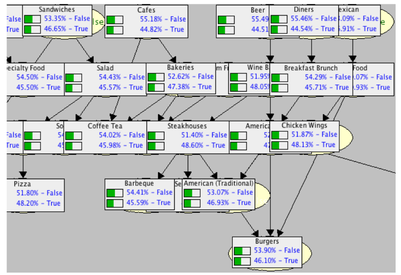
### *Using User Review Characteristics*

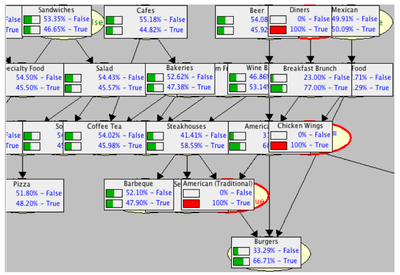
We also experimented with extracting adjectives from reviews and using this adjectives to model each business. The top one thousand adjectives were used to construct vectors representing each restaurant, where a restaurant's value for position 'X' in the vector is the number of times users described the restaurant with the adjective corresponding to position X. Using these features, our classifier missed by an average of 0.982 stars with the k-nearest neighbors.

## Bayesian Networks

To model the interactions between different categorial labels, we built Bayesian Networks to model an ontological relationship between various categories. Scores better than the restaurant average are considered "liking" the restaurant. We then build a Bayesian network by computing all pairwise independences and modeling statistically dependent relationships with Conditional Probability.

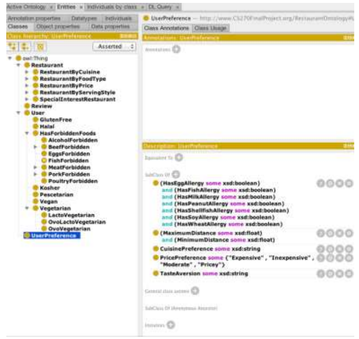
For example, we show how liking Diners, Chicken Wings and traditional American restaurants affects other categories

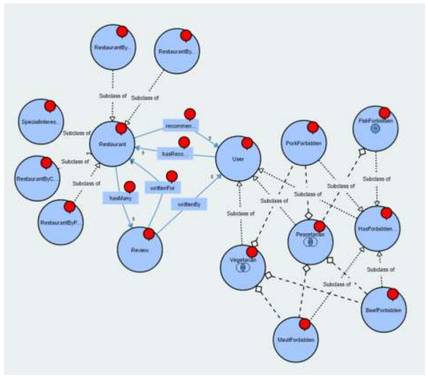



## Ontology

We constructed an ontology in Protege, which contains Restaurants, Reviews, Users, and User Preferences.





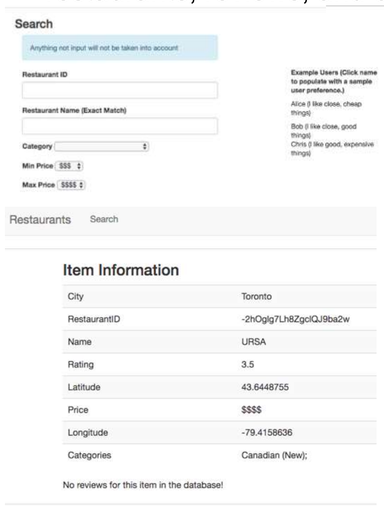

## Evaluation

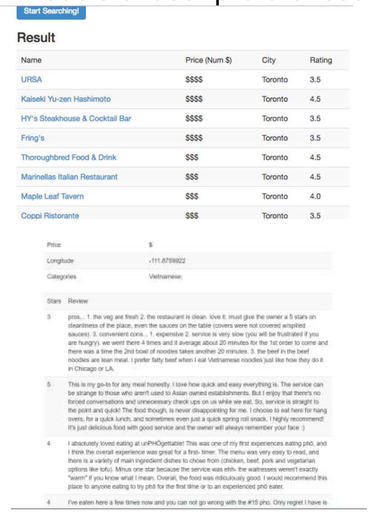
Users were asked qualitative questions about how useful the system was helpful addressing their information needs

- Our model requires users to specify a latitude and longitude for the "closest to" metric to be computed.
- Users didn't have any issues with navigating the UI.
- The search features were limited and very exact.

## SQL Web Application

We built a web front-end for searching and examining restaurants, reviews, and our models of user preferences



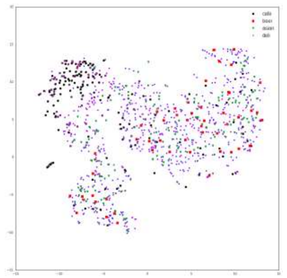



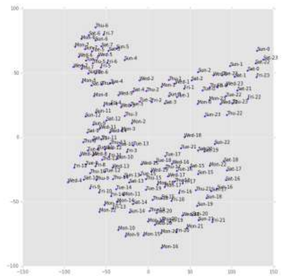




## Clustering & Similarity Measures

User checkin data was used to model, in an unsupervised way, the ontological similarity between attributes, such as business relationships or even the times of the week.





## Discussion

- The web application could be expanded to allow users to express more preferences, allowing us to return higher quality search results.
- Moreover, if we could acquire menu data, which would allow for more stringent filtering of foods & dietary restrictions.
- We found that the descriptors of a business found in user reviews were more useful in predicting a user's rating for a restaurant than yelp's given restaurant descriptors.