

# Yelp Restaurant Reviews Project Task Description

Qishen “Justin” Su, Kelley Lynch, Yuanyuan Ma

February 9, 2016

The goal of our project is to extract details about a restaurant’s menu items from Yelp reviews. Yelp reviews can be a valuable tool for consumers, who want to pick foods that suit their tastes, and restaurants, who can gain valuable feedback on their products. Reviews often feature a detailed description of many aspects of a customer’s experience at a restaurant, and while these reviews can be useful in choosing a restaurant, they make it difficult to decipher the quality of a restaurant’s food specifically. By annotating customer reviews and applying machine learning algorithms, we intend to extract information about specific qualities of food items from Yelp reviews.

Our corpus will consist of reviews provided by Yelp during round 7 of the annual Yelp Dataset Challenge. To limit the scope of this project, only reviews of Mexican restaurants with at least 100 reviews will be used. To further limit the scope of the annotation task, we will not include reviews in our corpus that do not mention any food items.

The annotation task will consist of marking specific food item names, ingredient names, and quality descriptions. Food items must be specific, for example, burrito is an acceptable food item, whereas the food in this restaurant is not. Ingredient names will be linked to the foods they are a part of. Quality tags will be placed on explicit descriptions of food quality. These tags will be linked to either a food or an ingredient and be labeled with positive, negative, or neutral. For example, in the sentence the beans in this burrito are tasty, burrito would be tagged as a food item, beans would be tagged as an ingredient and linked to burrito, and tasty would be linked to beans and labeled as positive.

By applying a machine learning algorithm to our annotated corpus we hope to develop a system to extract and sort positive and negative quality information about a restaurant’s food items. In addition, the hierarchical structure of the food and ingredient tags will allow us to identify specific ingredients that are influencing customer opinion about a given food.