

Final Project Proposal

By: Brian Carlo and Nicolas Soldi

1. Title of the project

FoodWise

2. List of the team members

Nicolas Soldi and Brian Carlo

3. Clear statement of project goals, including how part D (achievement of project goals) should be assessed

Create a system, possibly a web application, that will take in the name of a restaurant in San Francisco and an ingredient or dietary restriction (e.g. vegetarian vs. allergic to peanuts) and spot the presence of that ingredient/dietary restriction in any of the restaurant's menu items. The priority of objectives to achieve this goal would be the following:

1. Feature extraction
2. Clustering and analytics
3. Creating a taxonomy of ingredients, methods, dishes
4. Ingredient detection functionality
5. Recipe data supplements

The expectation is that an operational version of this application will have the capability to pick out which dishes don't meet a certain criteria. We would categorize the types of distinctions. The plan for proving the accuracy of this application would be to take a portion of the original data for testing purposes (it will not be included in the training set) and use the application to make predictions against it and have users verify the predictions. We are working toward automating this process, but at this point in time human verification is the best available option.

4. Clear statement of the roles of each team member will take on

Brian: Researching APIs, writing code to extract data from such APIs, and cleaning restaurant menu data and ingredients. Extracting features from menu item descriptions. Exploratory data analysis of the extracted features.

Nicolas: Feature extraction from recipes, tokenization of menu data and document clustering. Exploring the possibility of programmatically extending WordNet with extracted features from the recipes and menus. These would include cooking methods, ingredients and popular dish names.

5. List of resources, including data and software, that will be necessary for this project

Data Resources: Locu API and Factual API, both provide restaurant data. Yummly recipe data API. Resources

Software resources: Python, NLP libraries (likely NLTK and specifically Wordnet), Tableau, SciKit-Learn, Pandas, Numpy. For web application: python-friendly web framework such as Flask or Django.

6. Initial project plan, including week-by-week timing of work

Between the due date of this project and the final project presentation we have approximately four weeks to work on our objectives. Here's a week by week breakdown:

- Week 1 to Week 2: Data acquisition, cleaning and feature extraction
- Week 2 to Week 3: Initial clustering/analytics and taxonomy building
- Week 3 to Week 4: Analytics report generation and ingredient detection functionality
- Week 4 (optional): Supplementing model with recipe data