

Task 4 and 5 Overview

Overview

The general goal of Tasks 4 and 5 is to leverage recognized dish names to further help people making dining decisions. Specifically, Task 4 is to mine popular dishes in a cuisine that are liked by people; this can be very useful for people who would be interested in trying a cuisine that they might not be familiar with. Task 5 is to recommend restaurants to people who would like to have a particular dish or a certain type of dishes. This is directly useful to help people choose where to dine.

The two tasks are different, but they can be completed very similarly. Both tasks rely on knowledge about dish names. You are welcome to use the dish list provided based on the labeled annotations we have compiled, or you are free to use your own dish list based on your results from Task 3. In Task 4, your goal is to rank the dish names of a particular cuisine based on the reviews that mention these dish names so as to rank the popular dishes that received positive comments on the top. In Task 5, your goal is to rank the restaurants that offer a particular dish based on the comments about the dish expressed in the reviews of a restaurant so that the restaurants whose reviews contain many positive comments about the dish are ranked on the top.

Instructions

Some questions to consider when working on Tasks 4 & 5:

1. Given a cuisine and a set of candidate dish names of the cuisine, how do we quantify the popularity of a dish? How can we discover the popular dishes that are liked by many reviewers? What kind of dishes should be ranked higher in general if we are to recommend dishes of a cuisine for people to try? Would the number of times a dish is mentioned in all the reviews be a better indicator of a popular dish than the number of restaurants whose reviews mentioned the dish?

2. For people who are interested in a particular dish or a certain type of dishes, which restaurants should be recommended? How can we design a ranking function based on the reviews of the restaurants that mention the particular dish(es)? Should a restaurant with more dish name occurrences be ranked higher than one with more unique dish names?
3. How can you visualize the recommended dishes for a cuisine and the recommended restaurants for particular dishes to make them as useful as possible to users? How can the visualization be incorporated into a usable system? For example, you can imagine using the algorithms you developed for Tasks 4 and 5 to construct a system that allows a user to select a cuisine to see the favorite/popular dishes of the cuisine and further recommends the best restaurants if a user selects a particular dish or a set of dishes that are interesting to him/her.

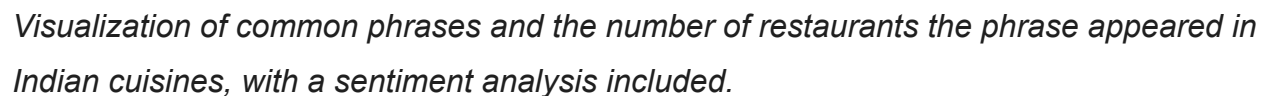
To receive credit, you must complete the following specific tasks. In general, it would be wise to first attempt to finish these required minimum tasks before exploring other ideas that you might have, especially if you have only limited time.

Task 4: Mining Popular Dishes

In this task, you will create a visualization showing a ranking of the dishes for a Yelp cuisine of your choice. You may use the [dish list we have provided](#), the list based on your annotations from Task 3 (or a subset of that list), or any other list for other cuisines. You might find it more interesting to work on a cuisine for which you can recognize many dishes than one with only a few dish names that you recognize.

There are many ways to approach this task; the main challenge will be how to create the ranking. You can devise your own method or use other methods you have learned in the Text Retrieval MOOC. The simplest approach can be to simply count how many times a dish is mentioned in all the reviews of restaurants of a particular cuisine, but you are encouraged to explore how to improve over this simple approach, e.g., by considering ratings of reviews or even sentiment of specific sentences that mention a dish. Even if you just try this simple approach, you may still need to consider options of

The following is a sample visualization created using d3.js. (You may click on the sample visualization to see a larger version in a new window.) You are encouraged to explore other visualization strategies. For example, an interesting visualization could show which dish names occur in what context – that is, they go beyond just the dish names.

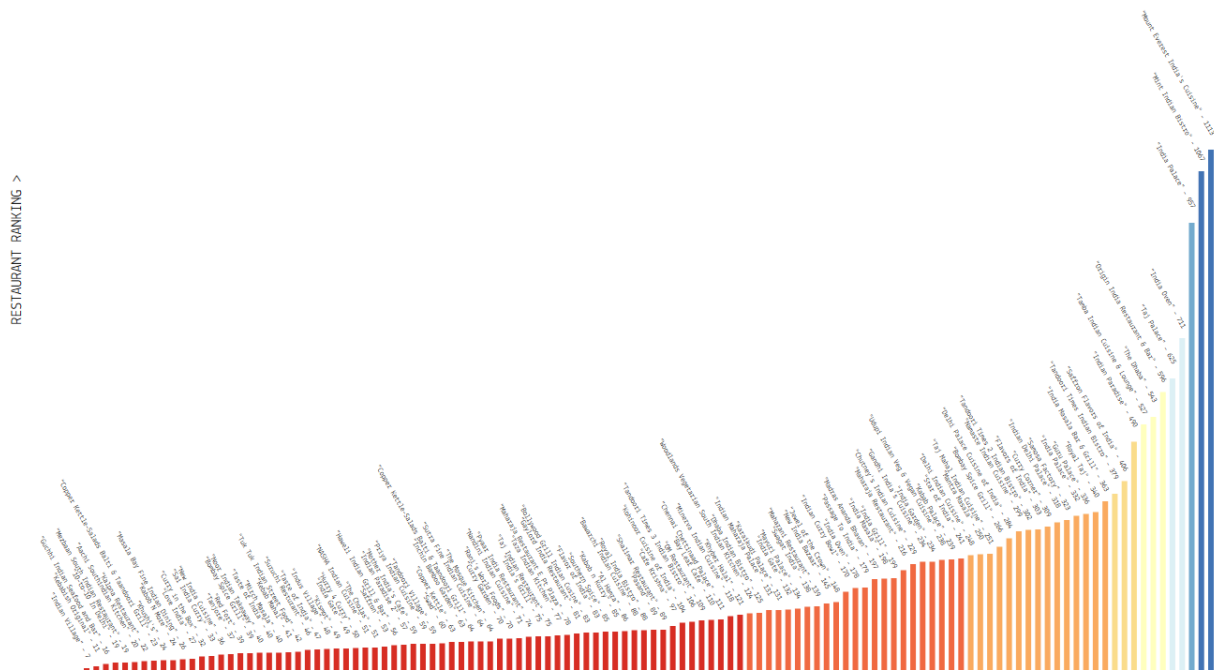


In this task, your goal is to recommend good restaurants to those who would like to try one or more dishes in a cuisine. Given a particular dish, the general idea of solving this problem is to assess whether a restaurant is good for this dish based on whether the reviews of a candidate restaurant have included many positive (and very few negative)

comments about the dish. You may choose a target dish or a set of target dishes from the list of "popular dishes" you generated from Task 4 or, otherwise, choose any dishes that have been mentioned many times in the review data (the more reviews you have for a dish, the more basis you will have for ranking restaurants).

You are required to create a visualization to show the ranking of the recommended restaurants. While a generic ranking of restaurants based on their overall ratings can be easily obtained, such a generic ranking is not as useful as one customized for a particular dish if one has decided to try this "particular dish." Thus, the ranking of restaurants you generated should be influenced somehow by the dish names you assumed to represent a diner's dining preference. The central question is thus how to design a dish-specific ranking algorithm for ranking restaurants. A simple approach easy to implement is to collect all the reviews mentioning a dish and compute the average ratings of these reviews for each restaurant so that a restaurant whose reviews containing the dish have the highest average rating would be ranked on the top. However, you are free to experiment with any parameters such as the rating of the restaurant, among other things.

Something to consider is to make your visualization general enough such that it could be used in a search engine or system and generate something useful for the users by recommending popular restaurants based on different dishes.



A visualization of the ranking of the restaurants in the Indian subset, based on the phrases mined.

Submission

You must submit a report in PDF format. We suggest that it be 2-3 pages long. Your report will need to include the following elements.

1. Two visualizations, one for each of the tasks
2. A written portion that provides sufficient detail for others to reproduce the process, including
 - A brief description of what you did, including any information retrieval techniques, how you devised your ranking function, the parameters you used, etc.
 - How you applied the models to the specific cuisine
 - Any interesting findings you made during your experience completing these tasks
 - Your opinions about whether the results you generated make sense or are useful in any way