CS 145 Milestone 2 - ToMEto

Jonathan Joo, Matthew Jin, Boyu (Charlie) Tong, Albert Ge

April 25, 2016

1 Goals

This week, we wanted to accomplish

1. Designing and planning algorithms to process the data.

2 Progress

We accomplished two preliminary algorithms this week.

2.1 Algorithms

The first algorithm (in `analyzer.py`) constructed a graph.

- Using individual ingredients as nodes, and then ranked the nodes by the number of recipes it appeared in with adjacent neighbors (ingredients in the same list).
- Each edge in the network also had an associated weight, which quantifies the number of recipes that the two connected ingredients both appear in.
- Obtained preliminary results on the most important ingredients in the network. Furthermore, used these ingredients to determine best-matching *complements* for a given recipe, which implements naively a ingredient that could be added into a recipe.
- For each ingredient in the recipe, we found their external ingredient (i.e. outside of the recipe), summing over the weights of each connection, normalizing by degree of each ingredient. Additionally, we ignored the top ten most common ingredients (so as to not suggest something obvious).

The second algorithm (in `analyzer_PMI.py`) constructs a simmilar graph, however instead of the weights being the number of recipes, we used the *Pointwise mutual information* metric, defined by

$$PMI(a,b) = \frac{\text{\# recipes containing a and b}}{(\text{\# recipes containing only a})(\text{\# recipes containing only b})}$$

2.2 Front-end

Additionally, we worked more on front-end.

- Got a few different pages up.
- Worked on logo design as well as more UI for the landing page, which can be seen in the "landing2" tab.
- Now able to handle input/output on search forms, which can now be integrated fully into our python scripts. Thus, using this implementation of our search, we can use search algorithms to find the proper recipes from our database to deliver to a user upon a search input.

2.3 Engineering Design

- Began refactoring our initial analyzer into the actual backend
- Brainstormed and discussed with Albert on algorithms

3 Contributions

Jonathan Joo: Front-end, begin linking to back-end

Matthew Jin: Reading on additional literature, helped Albert with choosing algorithms to try

Charlie Tong: Continued work on `analyzer.py`, implementing first algorithm

Albert Ge: Reading on additional literature, implemented PMI algorithm

4 Adjustments to plan

We are still in the stage of developing good algorithms for the network, which is inline with the M2 goals we had. We will probably continue to work on determining a good algorithm, and as we reach more optimal results, begin to seek out interviewers again for feedback on the results that we get.

Furthermore, as we are beginning to link front-end with back-end, we are thinking more about good design principles in engineering. We took a very sledgehammer approach to get things off the ground and running, however as we continue to grow our codebase, we feel that it will be important to refactor the code to make it more readable, scalable, and less bug-prone.