Brandon Dixon
Ken Jung
Adam Sudit
CMSC 122 Winter 2016

Project Proposal: aster*sk

A common conundrum inherent to the planning of weekend excursions is deciding where to eat. There are a number of factors that go into this decision: proximity to home, price, popularity on social media, favorable reviews in local publications or blogs, presence of a bar, type of fare being served, romantic ambiance, or even if a particular dish is on the menu. Furthermore, it is often difficult to check all of these decision parameters on the internet, since they are each located on different forums or sites. We hope that through our project we can fix this problem. Our group wants to create an application that, given a set of parameters that describe the desired restaurant(s), outputs a list of restaurants that come closest to the search criteria. Since we acknowledge that certain people may value some of the parameters more than others, the ordering of the restaurant listing would be determined based on the relative importance of each of the search parameters to the user.

There are a variety of data sources that would be useful in accomplishing our goal. For the proximity to home parameter, we would take advantage of the Google Maps API (thus the address of the user would also an important input). We could also use the Twitter API (among other social media API's) to find the number of times each restaurant in question is mentioned on Twitter. Finding menu related data probably presents the most interesting source of data for us. Ideally, the menu would be on a restaurant web page, but it is more often the case that the menu is a PDF/jpeg download. Thus we would have to find some sort of programming technology that would be able to convert PDF text into some sort of text file or string that can be more easily parsed. If the menu is on a webpage, we can simply use Beautiful Soup via the conventional web scraping method. Restaurant reviews could be found on sites like yelp, or the NY Times Restaurant section via the NY Times API. Thus, we see a rich variety of data sources that we could use for this project.

As mentioned previously, one potential programming technology that we would like to use is a PDF to text converter like the Phrase API (for Java). This would be important for making queries through certain restaurant menus. Since a lot of these API's are not for Python, SWIG would also be helpful so that we can use such API's with Python.

Timeline:

By end of fifth week: Come up with rating and characteristic-recognizing system for restaurants that can be pulled from the web, e.g. Twitter API / Google

By end of sixth week: Have a functioning search engine that can identify restaurants meeting the parameters

By end of seventh week: Generalize last two goals for other fun stuff, e.g. movies

By end of eighth week: Find a way to do the "planning" behind the night, i.e. making sure that events don't overlap, etc.

By end of ninth week: Make everything user-friendly and versatile