

CS122 Project Proposal

Title: Take Me Away!

Objectives

We want to develop a website that suggests affordable and well rated venues and attractions in the City of Chicago. This project is motivated by the need to provide fast access to places that match your preferences and have a good price-quality relationship. Sometimes people will take a lot of time reading over reviews or end up going to places they don't enjoy because information is spread out and overwhelming. We intend to make the choice easier by providing a tool that will select for you the top venues in a category and neighborhood of your choice.

From the front-end side, we will ask the user for information to refine the search criteria and return a map and a list with our main suggestions given the filters. The information we will be asking is:

- Neighborhood
- Type of establishment (restaurant, museum, shop)
- Key words to describe the sort of experience they want
- Target distance
- Time constraints

From the back-end side we will feed our website with information from YELP. We used YELP because it is the most comprehensive data source on venues with an available API (TripAdvisor did not allow us to access their API) We will connect to the YELP API and also scrape their website to obtain the complementary information the API does not provide.

We are aiming to obtain the following data:

- Name and location
- Type of establishment
- Price (YELP provides a \$ scale to determine price, i.e. \$, \$\$, \$\$\$)*
- Ratings (YELP provides a star scale to determine ratings)
- Reviews from users*
- Hours of operations

*Means scrapped data.

To prevent scrolling through the entire website we will use the search feature on YELP and crawl through the venues that have one or two dollar signs (\$ or \$\$) and are in a given neighborhood. Since the structure of a YELP search url follows a pattern, this task becomes easier to implement:

http://www.yelp.com/search?find_loc=Hyde+Park,+Chicago,+IL&start=0&attrs=RestaurantsPriceRange2.1

The previous link already has a structure that allows to crawl through the establishments in a neighborhood (find_loc=Hyde+Park,+Chicago,+IL) for a type of

establishment (attrs=Restaurants) with a given price range (PriceRange2.1). This way we will make more efficient our crawling through the website.

The information will be used to do a matching with the key words the user provided and the constraints on time, distance and neighborhood. The matching will be done through regular expressions and the findall function.

Innovative tools and databases

The project will rely on previous concepts learned in class such as crawling through a website, data structures and regular expressions. However, we will be using 2 concepts not covered in class:

1. Connecting to the API of YELP and querying information. We will gain access and implement a function that obtains the relevant information.
2. Using Google Maps to present information and obtain distance estimations. We will specifically be using the Google Maps Directions and Geocoding API.

If time permits, we will be interested in adding the following features (in order of priority):

1. Scrapping from other websites such as Timeout, Trip Advisor and blogs.
2. Recommend which neighborhoods to go based on a high price-quality concentration of establishments
3. Connect to the Foursquare of the user to obtain information on their preferences.

Deliverables

1. **Week 3 (Feb 9-12):** Database of venues, reviews, and events
2. **Week 5 (Feb 23-26):** Mapping function and natural language processing matching
3. **March 15:** Complete webpage

	Weeks						
	1	2	3	4	5	6	7
Connect to APIs	■	■	■	■	■	■	■
Store and process data through SQL	■	■	■	■	■	■	■
Select attractions given user restrictions	■	■	■	■	■	■	■
Scrape reviews	■	■	■	■	■	■	■
Natural language processing of reviews	■	■	■	■	■	■	■
Scrape events	■	■	■	■	■	■	■
Map schedule	■	■	■	■	■	■	■
Wrap it all together	■	■	■	■	■	■	■
Create front end	■	■	■	■	■	■	■