

Recipe Search

Yu-Cheng Huang
2018.06.27

Environment

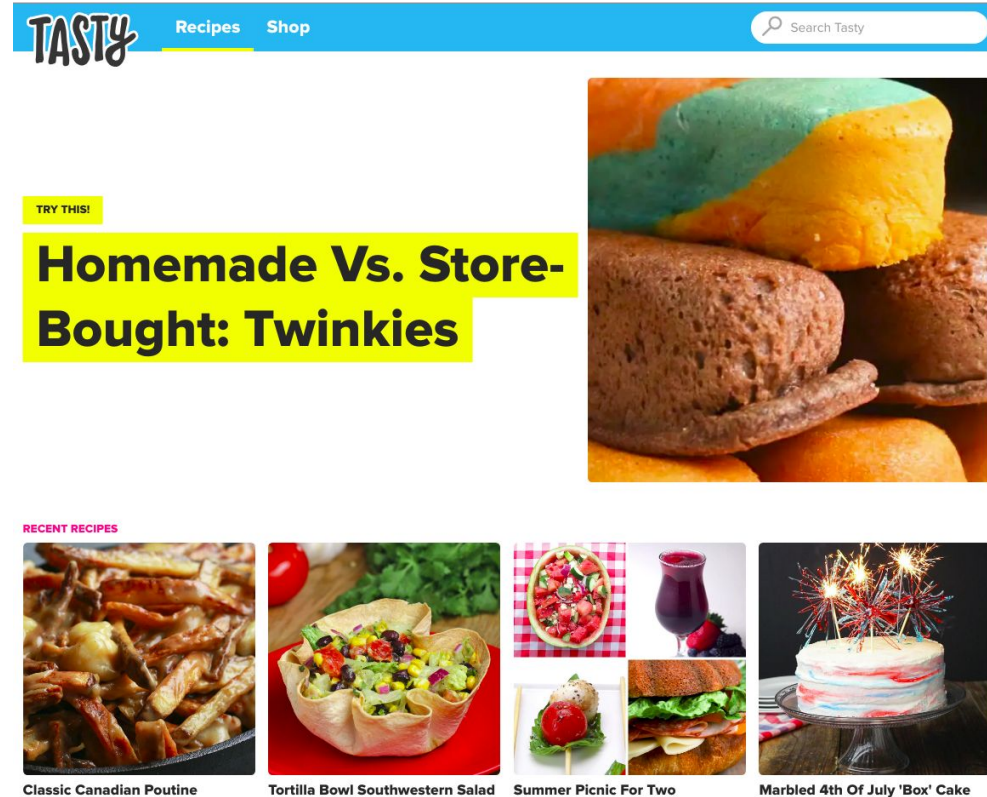
- Python 3.6
- Fedora 28
- Solr 7.3.0
- BeautifulSoup 4
- Jupyter, Tqdm, Matplotlib, Numpy, urllib
- Code: <https://github.com/amoshec/ccu-search-engine/tree/master/final>

Why

- It's often to see when you open the fridge and cannot determine what to cook.
- Recipe Search come to your rescue!
- Write down what ingredients you have as input, the application will search recipes for you.

Crawler

- Tasty
 - Popular Page on FB
 - 96M people following
 - Host their own site recently
 - <https://tasty.co/>



The screenshot shows the Tasty website interface. At the top is a blue navigation bar with the 'TASTY' logo, links for 'Recipes' and 'Shop', and a search bar. The main content area features a large article titled 'Homemade Vs. Store-Bought: Twinkies' with a yellow background. To the right of the article is a large image of Twinkies. Below the article, a 'RECENT RECIPES' section displays a grid of four recipe images: Classic Canadian Poutine, Tortilla Bowl Southwestern Salad, Summer Picnic For Two, and Marbled 4th Of July 'Box' Cake.

TASTY Recipes Shop

Search Tasty

TRY THIS!

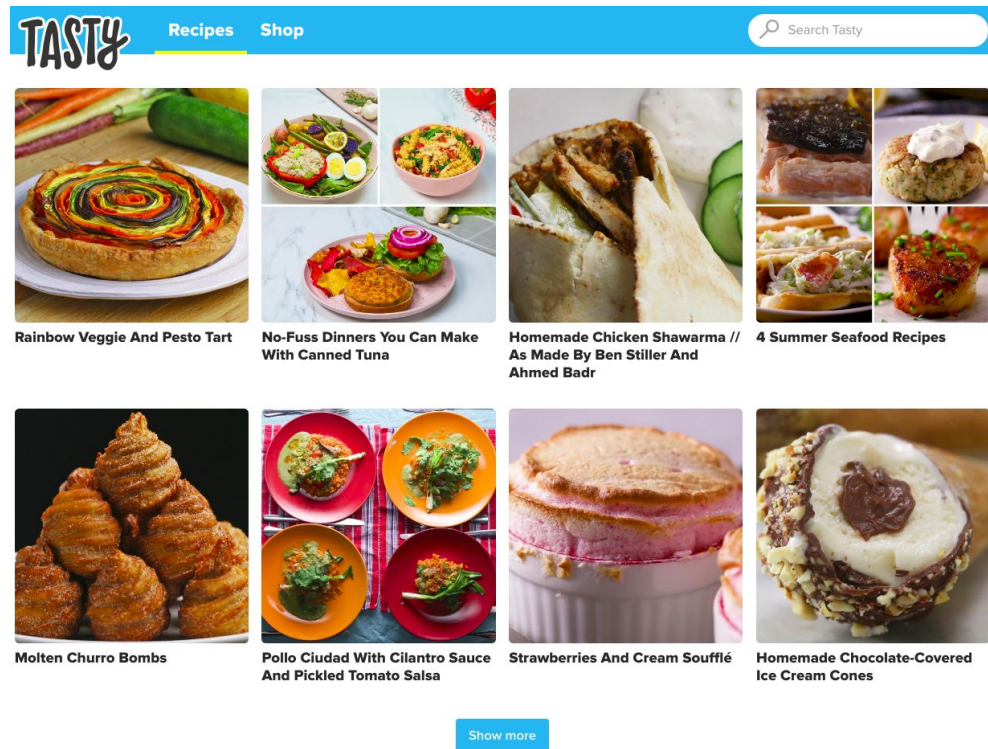
Homemade Vs. Store-Bought: Twinkies

RECENT RECIPES

- Classic Canadian Poutine
- Tortilla Bowl Southwestern Salad
- Summer Picnic For Two
- Marbled 4th Of July 'Box' Cake

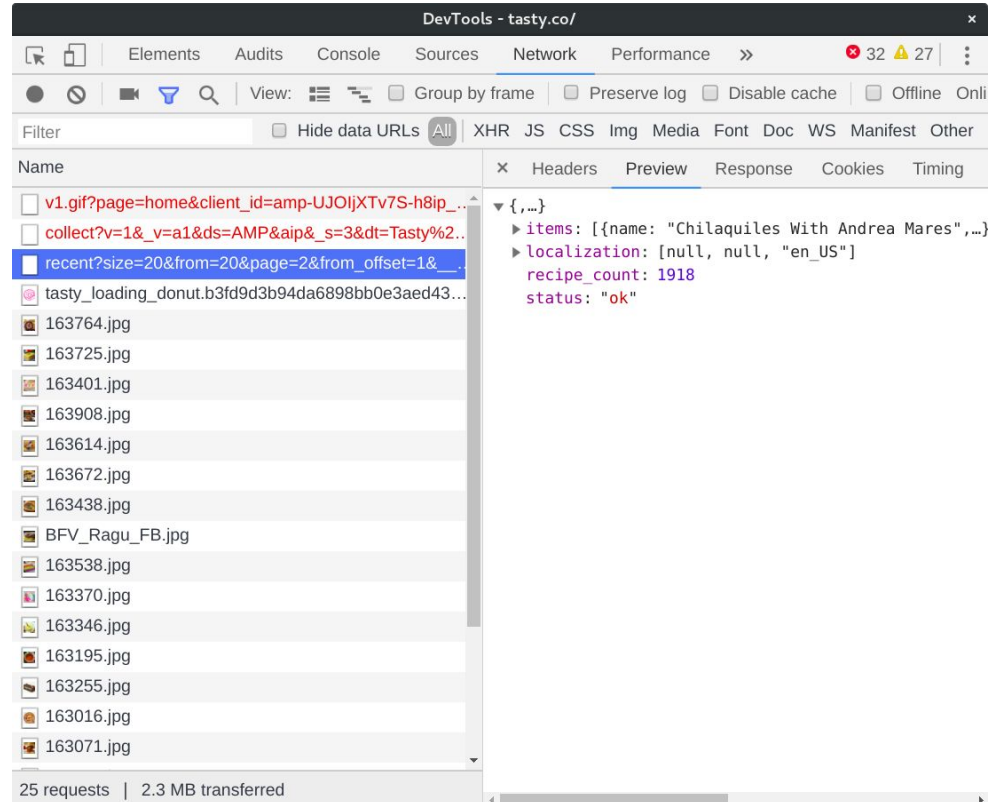
Crawler

- None of a page listing all recipes
- Recipe URLs can only be found when pressing “Show More” button
- Difficult to crawl



Crawler

- Chrome Dev Tool can monitor the requests of the page
- Found Tasty's API
- Returned value is JSON



Crawler

- Slugs Crawling
 - 06:32
 - 1266 recipes
 - 647 compilations
- HTML Crawling
 - 1:38:46
 - Average Ingredients: 15.82
 - Average Steps: 11.13

TASTY Recipes Shop

Crispy Country Fried Steak Roll

by Matthew Johnson

[Pinterest](#) [Facebook](#) [Email](#) [Print](#)


Ingredients

for 6 servings

- 1 tablespoon olive oil
- 1 onion, finely chopped
- 2 cloves garlic, minced
- 5 cups (200 g) spinach
- 2 lb (905 g) beef sirloin flap
- salt, to taste
- pepper, to taste
- 1 tablespoon garlic powder
- 1 tablespoon paprika
- 4 slices monterey jack cheese
- 2 cups (250 g) flour
- 5 eggs, beaten
- 3 cups (150 g) panko bread crumbs

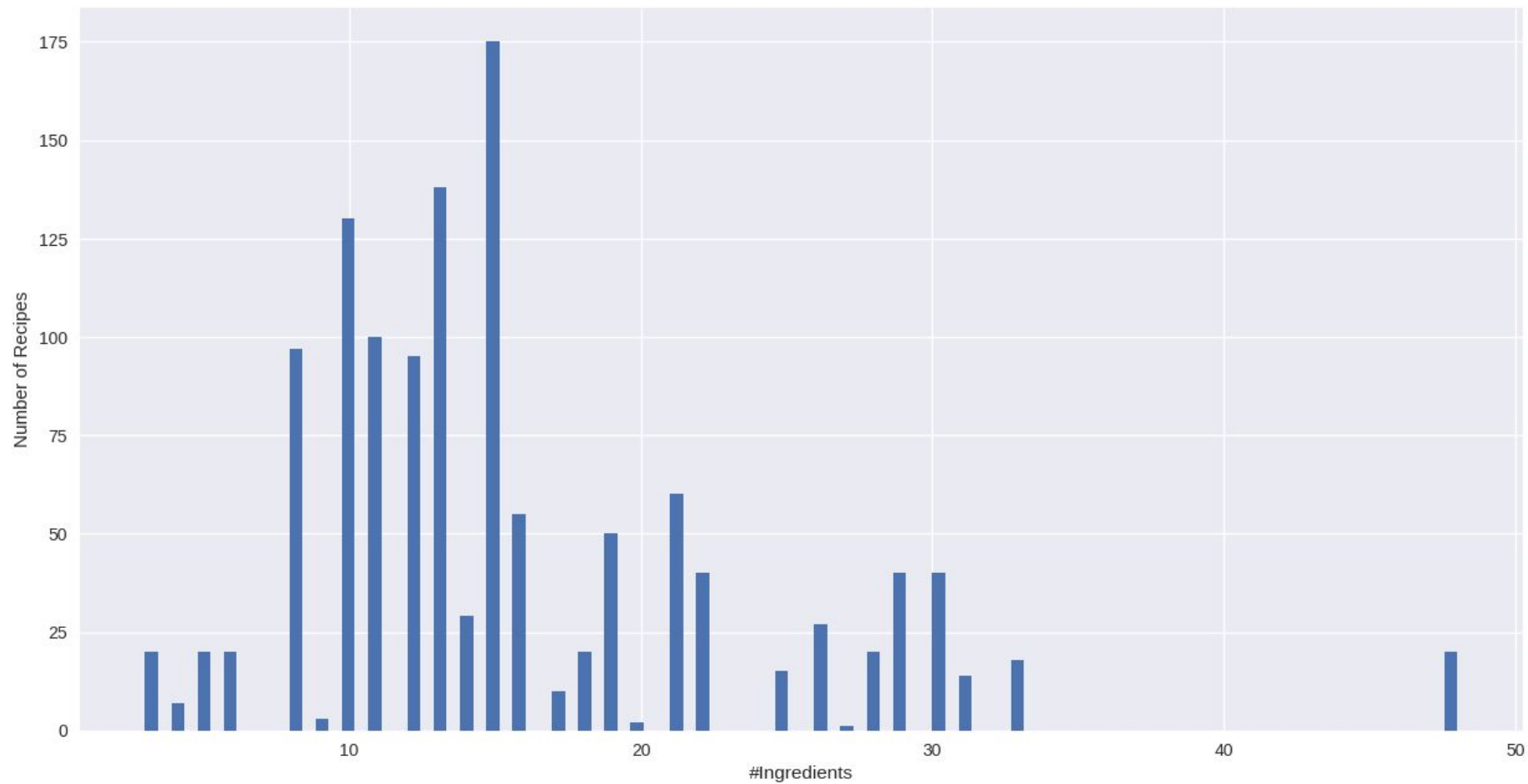
Preparation

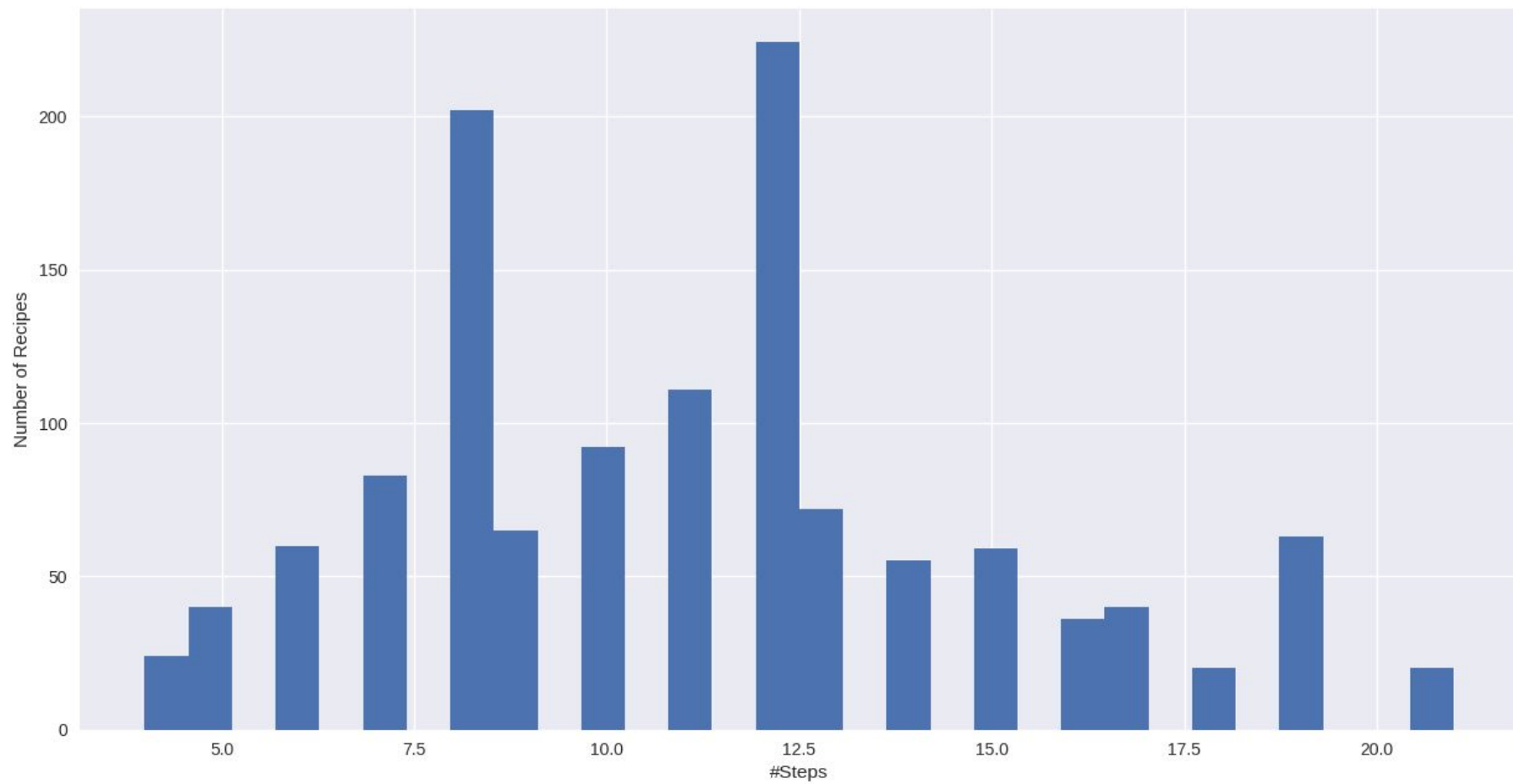
- 1 In a large pan, heat the olive oil over medium-high heat. Add the onion, garlic, salt, and pepper. Stir until softened, 3-4 minutes.
- 2 Add the spinach and stir until wilted down. Remove the pan from the heat.
- 3 Season the steak on all sides with salt, pepper, garlic powder, and paprika. Lay a piece of parchment paper over the cubed steak. Using a meat mallet or rolling pin, pound the steak until ¼-inch (6-mm) in thickness.
- 4 Remove the parchment paper. Lay the slices on Monterey Jack evenly over the cube steak. Add the sauteed spinach mixture over the cheese.
- 5 Gently roll up the steak, making sure to keep it as tightly wrapped as possible. Place the steak log on a sheet of plastic wrap. Wrap the steak up, tightly securing each side.



sirloin flap steak
2 lb

Inspired by [tasty.co](#)





Search Engine

- Solr is used
- Recipe {
 - slug: str
 - title: str
 - ingr: []
 - step: []
- }

Problems

- Some common ingredients is written in recipe, but won't be key in when searching.
 - water,
 - salt,
 - oliver oil,
 - etc
- Searched recipe use more ingredients than what you key in.
 - What you have: lettuce, onion, tomato, pasta
 - What recipe use: lettuce, onion, tomato, pasta, **shrimp**
 - But you want the recipe use only the ingredient you had.
 - Or sometimes you don't mind buy additional ingredient.

Problems

- 2 problems are solved by same method: introducing the ``n_ingr`` property.
- Recipe {
 - slug: str
 - title: str
 - **n_ingr: int,**
 - ingr: []
 - step: []
- }

Problems

- Sorted all candidate recipes by the distance of `n_ingr` and the number of ingredients you have
 - if you have 4 ingredients,
 - recipes with 4 ingredients appear first,
 - followed by recipes with 5 or 6 ingredients
 - ...
- Not a good enough solution, but it should work to some extent.

Problems

- **Solr doesn't support such sorting method!!!**
 - At least I didn't find it
- **Solution: Manual Hacking**
 - Ingredient: lettuce, onion, tomato, pasta
 - ↓
 - (ingr:lettuce AND ingr:onion AND ingr:tomato AND ingr:pasta)
 - (n_ingr:[2 TO 6])
 - $(n_ingr:2)^{0.5} (n_ingr:3)^{0.8} (n_ingr:4)^{1.5} (n_ingr:5)^{0.8} (n_ingr:6)^{0.5}$
- Result is not quite good, so I fallback to use pandas instead

Result

- Using TinyDB
 - Able to perform exact match
 - Dataset is small
 - Easy to use, Python XD

```
inp = ['bread', 'butter', 'chocolate', 'ice cream']
def use_only(ingrs):
    return all(any(x in text for x in inp) for text in ingrs)
res = db.get(where('ingr').test(use_only))
print(res['title'])
```

Ice Cream Sandwich Swirl Cake Recipe by Tasty