Recipe Search

Yu-Cheng Huang
2018.06.27

Environment

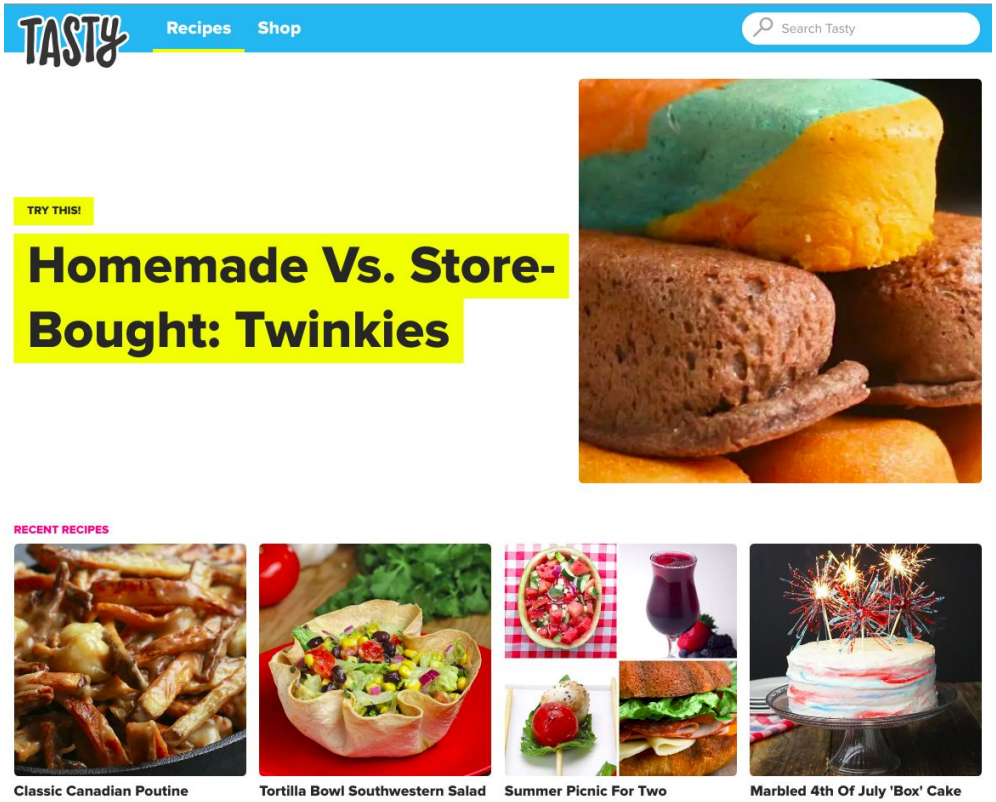
- Python 3.6
- Fedora 28
- Solr 7.3.0
- BeautifulSoup 4
- Jupyter, Tqdm, Matplotlib, Numpy, urllib
- Code: https://github.com/amoshyc/ccu-search-engine/tree/master/final

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.

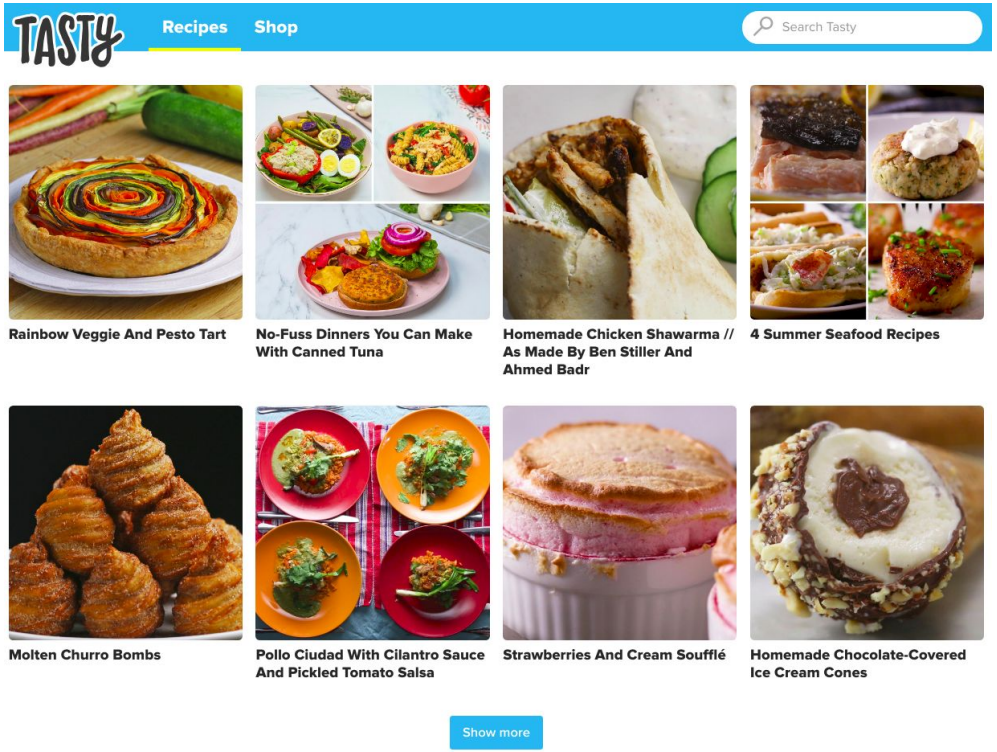
Crawler

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



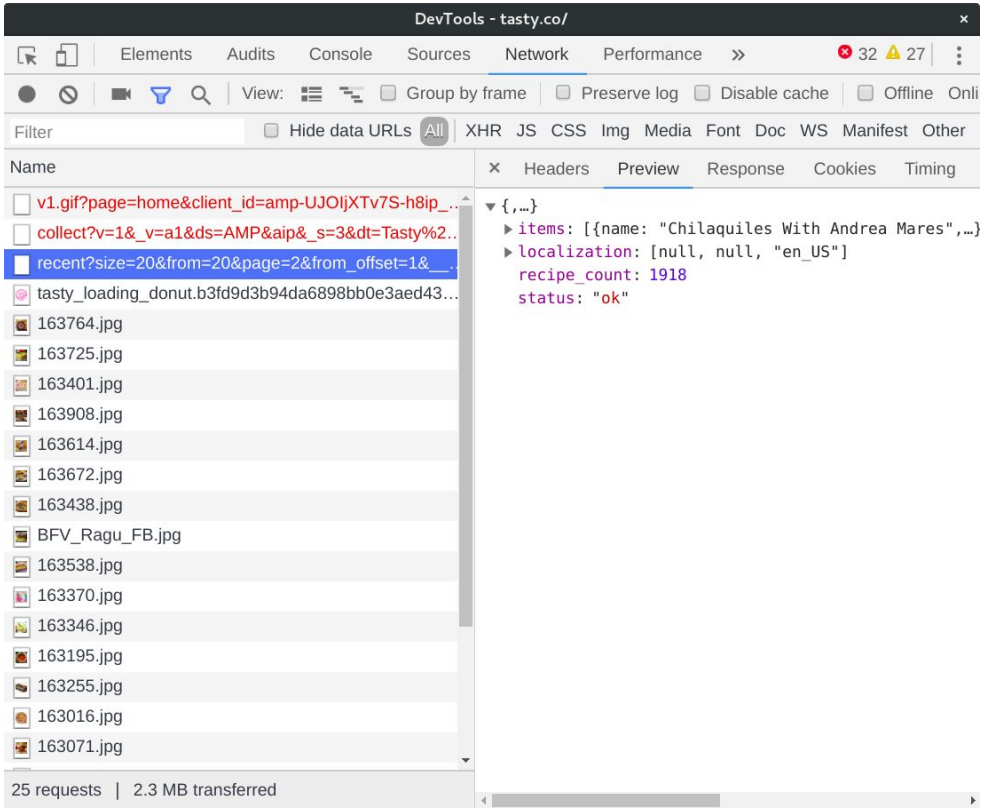
Crawler

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



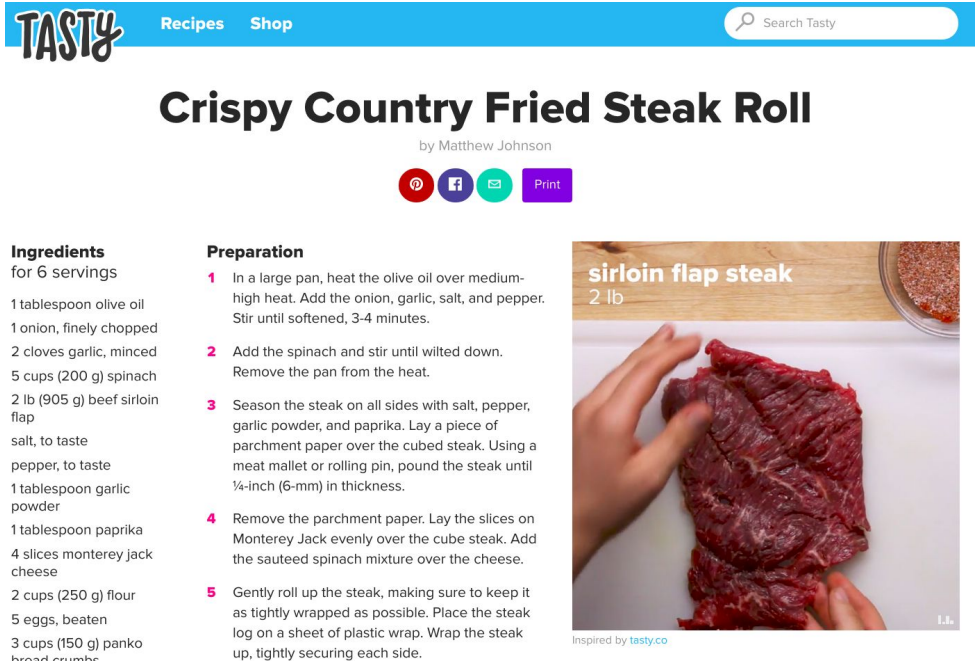
Crawler

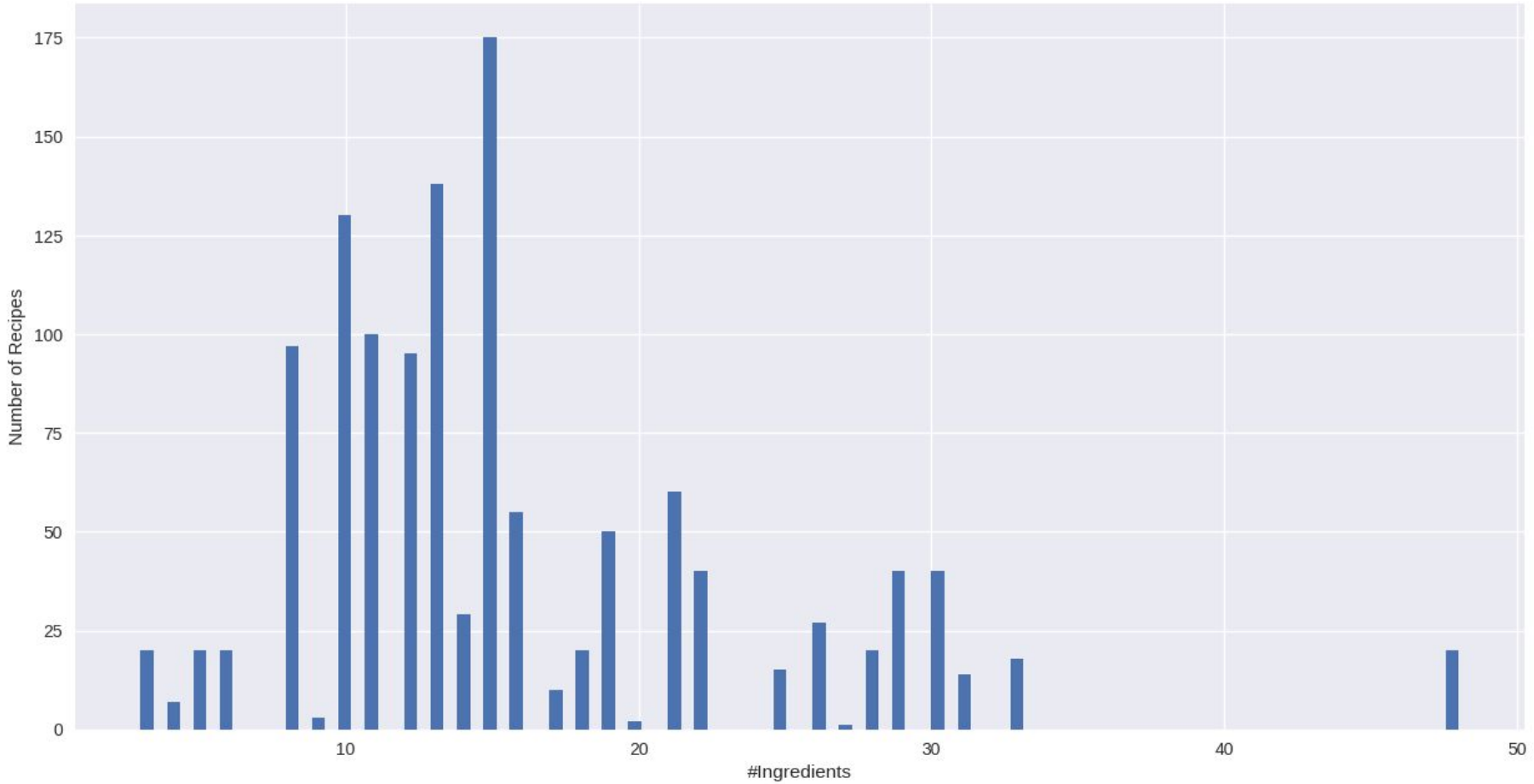
- Chrome Dev Tool can monitor the requests of the page
- Found Tasty's API
- Returned value is JSON

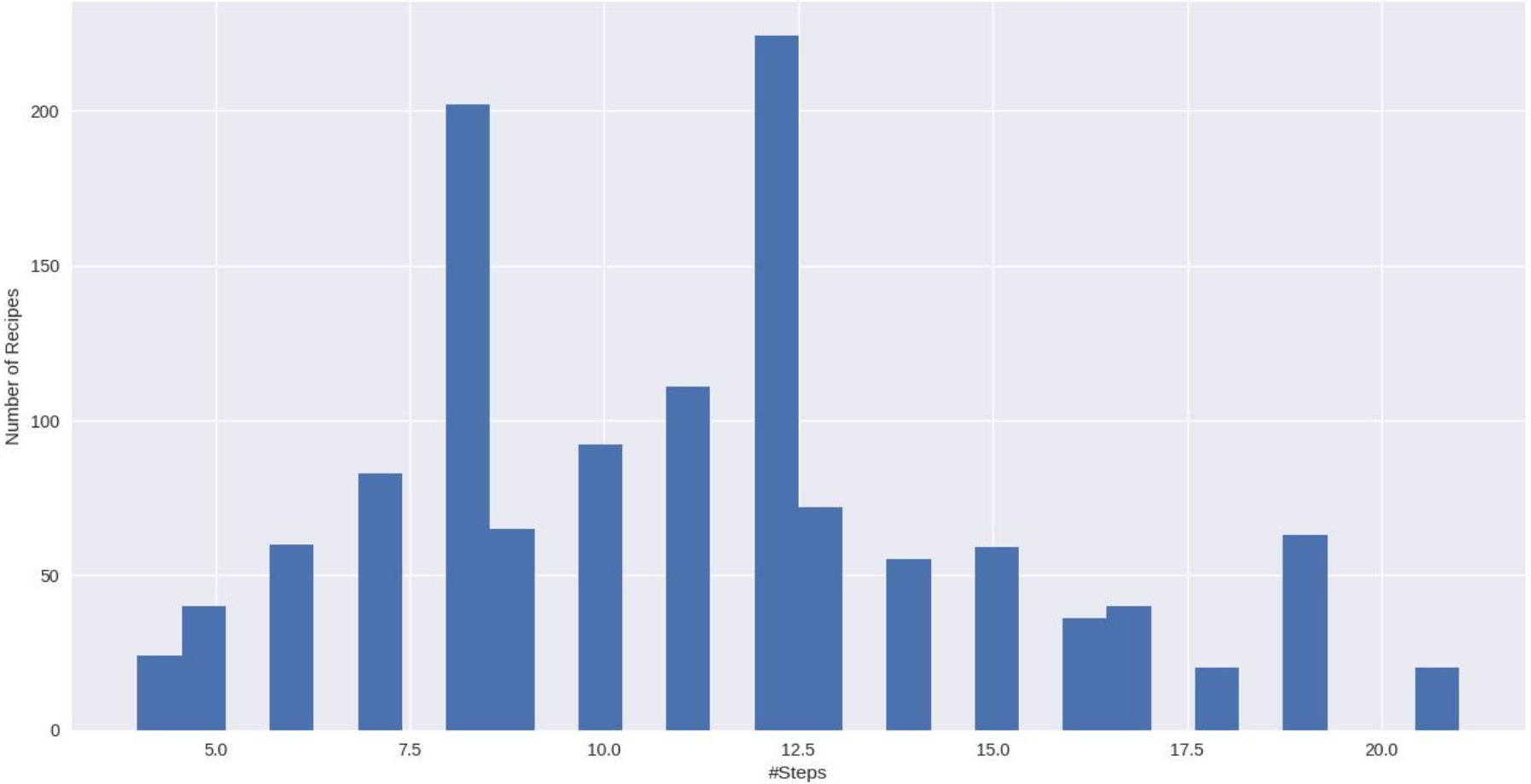


Crawler

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




175
150
125
100
75
50
25
0
Number of Recipes
10
20
30
40
50
#Ingredients


Number of Recipes
0
50
100
150
200
5.0
7.5
10.0
12.5
15.0
17.5
20.0
#Steps

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 recipse 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 candiadate 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: Manunal 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
```