

Yummly Web Crawler

Web crawler to download recipes from Yummly in JSON format

This is a web crawler and web scraper designed to download recipes from Yummly in the following format:

```
[
{
  recipeUrl: string,
  recipeName: string,
  recipePhoto: string,
  ingredients: string[],
  ratings: float,
  cookTime: string,
  serve: number
  tags: string[]
  recipeType: enum ( main course, drink, breakfast, desert, salad)
}
...
]
```

Getting started

Install Python 3.7 and Virtualenv

- Python 3.7 using pip
- Virtualenv using pip install virtualenv

Create virtual environment and activate it

Create a virtual environment using Python 3.7

```
python3.7 -m virtualenv venv
```

Activate the virtual environment

```
source venv/bin/activate
```

Install dependencies

Download all the dependencies to run the program using the given requirements.txt file.

```
pip install -r /path/to/requirements.txt
```

How to use

Navigate to the top level directory with the scrapy.cfg file

```
yummly
  venv
  yummlly_spider
    README.md
    scrapy.cfg
    yummlly_spider
```

Run using the following command:

```
scrapy crawl recipes
```

Press Ctrl-C to stop the crawler anytime

Check `output.txt` to see the crawled data