Scrapy Cheatsheet

This cheatsheet provides a quick reference for the key features of Scrapy, a Python web crawling and web scraping framework. Use this cheatsheet as a reference to help you write Scrapy code more efficiently.

Installation

You can install Scrapy using pip:

```
pip install scrapy
```

Creating a new Scrapy project

```
scrapy startproject project_name
```

Spiders

Creating a new spider

```
scrapy genspider spider_name domain.com
```

Defining a spider

```python
import scrapy

class SpiderName(scrapy.Spider):
    name = 'spider_name'
    start_urls = ['https://domain.com']

    def parse(self, response):
        # Code to extract data from the response
```

Items

Defining an item

```python
import scrapy

class ItemName(scrapy.Item):
    field1 = scrapy.Field()
    field2 = scrapy.Field()
```

Yielding an item

```python
yield ItemName(field1=value1, field2=value2)
```

Pipelines

Defining a pipeline

```python
class PipelineName:
    def process_item(self, item, spider):
        # Code to process the item
        return item
```

Enabling a pipeline

```python
ITEM_PIPELINES = {
    'project_name.pipelines.PipelineName': 300,
}
```

Settings

Common settings

```python
BOT_NAME = 'project_name'
ROBOTSTXT_OBEY = True
USER_AGENT = 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML,
like Gecko) Chrome/58.0.3029.110 Safari/537.36'
```

Custom settings

```python
DOWNLOAD_DELAY = 3
CONCURRENT_REQUESTS = 1
```

Running a spider

```
scrapy crawl spider_name
```

Resources

- [Scrapy documentation](#)
- [Scrapy tutorial](#)
- [Scrapy shell cheatsheet](#)