

# WEB SCRAPER

**Using scrapy**

**Web Scraping** (also termed Screen **Scraping**, **Web** Data Extraction, **Web** Harvesting etc.) is a technique employed to extract large amounts of data from websites whereby the data is extracted and saved to to a database or in a spreadsheet file/json file.

# SCRAPY

**Scrapy** is a free and open source web crawling framework, written in Python. Originally designed for web scraping, it can also be used to extract data using APIs or as a general purpose web crawler.

You can define “Rules” in scrapy on the basis of which scrapy will download only those web pages.

Once you have response, then the webpage can be parsed and data can be extracted.

# WHO IS USING SCRAPY TO DO BUSINESS ....

## SHOULD YOU BE ON THIS LIST?

<http://usedaywatch.com/> -uses Scrapy to empower real-time business information retrieval from Daily Deal sites.

<https://www.parsely.com/data-analysts/> -Uses Scrapy to scrape articles from hundreds of news sites.

<https://www.lyst.com/> - Uses Scrapy to crawl and scrape the fashion websites they index.

WANNA TRY URSELF ....

# INSTALLATION

1. Install Scrapy inside a virtual environment to avoid conflicting with your system packages.

```
--$ [sudo] pip install virtualenv
```

```
--$ pip install Scrapy
```

EXAMPLE -

```
import scrapy

class QuotesSpider(scrapy.Spider):

    name = "quotes"

    start_urls = [

        'http://quotes.toscrape.com/page/1/', ]

    def parse(self, response):

        for quote in response.css('div.quote'):

            yield {

                'text': quote.css('span.text::text').extract_first(),

                'author': quote.css('span small::text').extract_first(),

                'tags': quote.css('div.tags a.tag::text').extract(),

            }
```



**Thank you**