BASLIK 1

BASLIK 2

BASLIK 3

BASLIK 4

To install Scrapy using conda, run:

conda install -c conda-forge scrapy

Alternatively, if you’re already familiar with installation of Python packages, you can install Scrapy and its dependencies from PyPI with:

pip install Scrapy

* Komut sistemini acmak

scrapy shell

* Proje olusturmak

scrapy startproject projectName

LINKLERI YAKALAMAK

*import* scrapy

class QuotesSpider(scrapy.Spider):

    name = "quotes" #*name onemli, unique olmali, cmd den cagrilacak*

    def *start\_requests*(self):

        urls = [

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

            'http://quotes.toscrape.com/page/2/',

        ] #*2 tane web sitesine gidecegimizi soylutor*

*for* url *in* urls:

*yield* scrapy.Request(url=url, callback=self.parse)

    def *parse*(self, response):

        page = response.url.split("/")[-2]

        filename = 'quotes-%s.html' % page

*with* open(filename, 'wb') *as* f:

            f.write(response.body)

        self.log('Saved file %s' % filename)

#*title alaini alir request yapilan yerdeki*

#*response.css('title').extract() #\* ['<title>Quotes to Scrape</title>']*

#*response.xpath('//title') #\* ['<title>Quotes to Scrape</title>']*

#*response.css('title::text').extract() #\* ['Quotes to Scrape']*

#*response.css('title::text').extract()[0] #\*'Quotes to Scrape'*

#*response.css('title::text').extract\_first() #\*'Quotes to Scrape'*

#*response.xpath('//title/text').extract\_first() #\*'Quotes to Scrape'*

#*Sayfadan classi quote olan divlerden 0. yani ilk siradakini secmek*

quote = response.css("div.quote")[0]

#*daha sonra bu quote icindeki classi text olan spani secmek ve degiskene atamak*

title = quote.css("span.text").extract\_first()

#*daha sonra title diyerek spanimizi yazdirabiliriz terminalde*

#*classi author olan small in textini al ve donen arrayden extract first ile ilk siradakini al veya arrayden alir gibi cek*

author = quote.css("small.author::text").extract\_first()

author = quote.css("small.author::text").extract()[0]

#*MULTIPLE SELECTOR*

#*classi tags olan divi al ve onun icindeki class i tag olan a larin textini al*

tags = quote.css("div.tags a.tag::text").extract()

VERILERI JSON’A YAZDIRMAK

* JSON’a cikti almak icin kodlarimizi asagidaki gibi yazip terminale asagidaki calistir kodumuzu yazdiktan sonra output anlamindaki -o isareti ve json dosya adini yazip JSON ciktimizi alabiliriz.

**scrapy crawl quotes -o quotes.json**

*import* scrapy

class QuotesSpider(scrapy.Spider):

    name = "quotes" #*name onemli, unique olmali, cmd den cagrilacak*

    start\_urls = [

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

      'http://quotes.toscrape.com/page/2/',

    ]

    def *parse*(self, response):

      #*!JSONA YAZDIRMAK ICIN*

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

        title = quote.css('span.text::text').extract()[0]

        author = quote.css('small.author::text').extract()[0]

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

*yield* {

          "title": title,

          "author": author,

          "tags": tags

        }

VERILERI TXT DOSYASINA YAZDIRMAK

* Asagidaki gibi dongumuzu kurduktan sonra asagidaki kodla scrappy’i calistiririz.

**scrapy crawl quotes**

*import* scrapy

class QuotesSpider(scrapy.Spider):

    name = "quotes" #*name onemli, unique olmali, cmd den cagrilacak*

    start\_urls = [

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

      'http://quotes.toscrape.com/page/2/',

    ]

      #*!TXT DOSYASINA YAZDIRMAK ICIN*

*with* open('quotes.txt', 'w', encoding='utf-8') *as* file:

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

          title = quote.css('span.text::text').extract()[0]

          author = quote.css('small.author::text').extract()[0]

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

          file.write('\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\n')

          file.write('Quote : ' + title + '\n')

          file.write('Author : ' + author + '\n')

          file.write('Tags : ' + str(author) + '\n')