import scrapy

from scrapy import Request

class JobsSpider(scrapy.Spider):

name = "jobs"

allowed\_domains = ["craigslist.org"]

start\_urls = ["https://newyork.craigslist.org/search/egr"]

def parse(self, response):

jobs = response.xpath('//p[@class="result-info"]')

for job in jobs:

relative\_url = job.xpath('a/@href').extract\_first()

absolute\_url = response.urljoin(relative\_url)

title = job.xpath('a/text()').extract\_first()

address = job.xpath('span[@class="result-meta"]/span[@class="result-hood"]/text()').extract\_first("")[2:-1]

yield Request(absolute\_url, callback=self.parse\_page, meta={'URL': absolute\_url, 'Title': title, 'Address':address})

relative\_next\_url = response.xpath('//a[@class="button next"]/@href').extract\_first()

absolute\_next\_url = "https://newyork.craigslist.org" + relative\_next\_url

yield Request(absolute\_next\_url, callback=self.parse)

def parse\_page(self, response):

url = response.meta.get('URL')

title = response.meta.get('Title')

address = response.meta.get('Address')

description = "".join(line for line in response.xpath('//\*[@id="postingbody"]/text()').extract())

compensation = response.xpath('//p[@class="attrgroup"]/span[1]/b/text()').extract\_first()

employment\_type = response.xpath('//p[@class="attrgroup"]/span[2]/b/text()').extract\_first()

yield{'URL': url, 'Title': title, 'Address':address, 'Description':description, 'Compensation':compensation, 'Employment Type':employment\_type}