import scrapy

from scrapy import Request

class CrawlerSpider(scrapy.Spider):

name = 'crawler'

allowed\_domains = ['https://seattle.craigslist.org/search/est/cta','https://seattle.craigslist.org']

start\_urls = ['https://seattle.craigslist.org/search/est/cta/']

def parse(self, response):

#Getting all ads on a single page

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

for result in results:

title = result.xpath('h3[@class="result-heading"]/a[@class="result-title hdrlnk"]/text()').extract\_first()

price = result.xpath('span[@class="result-meta"]/span[@class="result-price"]/text()').extract\_first()

neighbourhood = result.xpath('span[@class="result-meta"]/span[@class="result-hood"]/text()').extract\_first().strip('+)').replace("(","")

url = result.xpath('h3[@class="result-heading"]/a/@href').extract\_first()

#Browsing to the ad page and calling the parse\_page function to scrape the description

yield Request(url, callback = self.parse\_page, dont\_filter=True,

meta={'title':title, 'price':price, 'neighbourhood': neighbourhood, 'url': url})

#Getting the link for the next button

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

next\_url = "https://seattle.craigslist.org" + next\_button

yield Request(next\_url, callback=self.parse, dont\_filter=True)

def parse\_page(self, response):

title = response.meta['title']

price = response.meta['price']

neighbourhood = response.meta['neighbourhood']

url = response.meta['url']

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

yield {'title':title, 'price':price, 'neighbourhood': neighbourhood, 'url': url, 'ad\_description': ad\_content}

