Kuandyk Miras

# -\*- coding: utf-8 -\*-

import scrapy

import requests

class KrishaSpider(scrapy.Spider):

name = 'krisha'

allowed\_domains = ['krisha.kz']

start\_urls = ['http://krisha.kz/pro/specialist/']

initial\_next\_urls = []

page = 0

def parse(self, response):

self.page += 1

with open('krisha.csv', 'w', encoding='utf-8') as f:

realtors = response.xpath("//div[@class='pitem']")

for i in realtors:

#photo = i.css('div.image\_cover').css('img.photo/@href').extract\_first()

name = i.css('div.pr\_block').css('a.name::text').extract\_first()

spec = i.css('div.pr\_block').css('div.speciality::text').extract\_first()

numbers = i.css('div.pr\_block').css('div.phones::text').extract\_first().replace(' ', '').replace('(', '').replace(')', '').split(',')

# photo = i.css('div.image\_cover').extract\_first()

# name = i.css('div.pr\_block').css('a.name::text').extract\_first()

# numbers = i.css('div.pr\_block').extract\_first().replace(' ', '').replace('(', '').replace(')', '').split(',')

f.write(u'{},{}\n,{}'.format(name, spec, numbers))

next\_page\_url = response.xpath("//link[@id='NextLink']/@href").extract\_first()

if next\_page\_url and next\_page\_url not in self.initial\_next\_urls:

if len(self.initial\_next\_urls) < 2:

self.initial\_next\_urls.append(next\_page\_url)

yield scrapy.Request(response.urljoin(next\_page\_url))