```
In [ ]: #!/usr/bin/env python
# coding: utf-8
# import of relevant libraries
import scrapy
from scrapy.crawler import CrawlerProcess
from scrapy import Selector
from scrapy.linkextractors import LinkExtractor
from fiatchrysler_scraper.items import FiatChryslerItem
import numpy as np
from scrapy.loader import ItemLoader
import csv
from datetime import datetime
#creating spider for Fiat Chrysler
class FinanceNewsScraperSpider(scrapy.Spider):
    name = "fiatchryslernewsarticles"
    def start requests(self):
        start_urls = ['https://www.reuters.com/companies/FCHA.MI/news',
        urls = start_urls
        for url in urls:
            yield scrapy.Request(url=url, callback=self.parse newspage)
    def parse newspage(self, response):
        links = response.xpath('//a[contains(@href,"/article/")]/@href').extract()
#extract hyperlink
        for url in links:
            yield response.follow(url = url,callback = self.parse article)
    def parse article(self, response):
        item = FiatChryslerItem()
        item['article link'] = response.url
        item['article headline'] = response.xpath('//*[contains(@class,"ArticleHead
er headline")]/text()').extract()
        item['article date'] = response.xpath('//*[contains(@class,"ArticleHeader d
ate")]/text()').extract()
        item['article text'] = response.xpath('//div[@class="StandardArticleBody bo
dy"]//p/text()').extract()
        print(item)
        #saving data to file.
        path = 'news/'
        file = 'fiatchrylsernews_' + str(datetime.now().strftime("%Y%m%d-%H%M")) +
'.csv'
        file name = open(path + file, 'a')
        fieldnames = ['article_link', 'article_headline','article_date','article_te
xt'] #adding header to file
        writer = csv.writer(file name, lineterminator='\n')
        writer.writerow([item[key] for key in item.keys()])
```

1 von 1 25.09.2020, 13:38