In [ ]:

```
#!/usr/bin/env python
# coding: utf-8

# import of relevant libraries
import scrapy
from scrapy.spiders import CrawlSpider
from scrapy.crawler import CrawlerProcess
from scrapy import Selector
from scrapy.linkextractors import LinkExtractor
from audi_scraper.items import AudiItem
import numpy as np
from scrapy.loader import ItemLoader
import csv
from datetime import datetime
from scrapy_splash import SplashRequest
from selenium import webdriver
from scrapy import FormRequest

#creating spider for Audi

class FinanceNewsScraperSpider(scrapy.Spider):
    name = "audinewsarticles"

    #def __init__(self):
    #    self.driver = webdriver.Firefox(executable_path=r'D:\victo\Documents\Frankfurt School\Master Thesis\geckodriver.exe')

    def start_requests(self):
        start_urls = ['https://www.reuters.com/companies/NSUG.DE/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 scrapy.Request(url=url,
                                 # meta={'splash':{'args':{'html': 1,
                                 #                         'wait': 0.5,
                                 #                         'har': 1
                                 #                         },
                                 #                 'endpoint': 'render.json'
                                 #                 }
                                 #       },
                                 callback=self.parse_article)

    def parse_article(self, response):
        item = AudiItem()
        item['article_link'] = response.url
        item['article_headline'] = response.xpath('//*[contains(@class,"ArticleHeader_headline")]/text()').extract()
        item['article_date'] = response.xpath('//*[contains(@class,"ArticleHeader_date")]/text()').extract()
        item['article_text'] = response.xpath('//div[@class="StandardArticleBody_body"]//p/text()').extract()

        print(item)
```

```
        #saving data to file.
        path = 'news/'
        file = 'audinews_' + 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()])
```