2.获取数据

简单的获取网页，网页文本

```
response = requests.get(url).text
```

对于很多网站可能需要用户身份登录，此时用headers伪装，此内容可以在浏览器f12获得

```
headers = {
    'Cookie': 'cookie，非真实的',
    'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64)
AppleWebKit/537.36 (KHTML, like Gecko) Chrome/125.0.0.0
Safari/537.36'
}

headers = {
    'Host': 'www.qidian.com',
    'Connection': 'keep-alive',
    'Pragma': 'no-cache',
    'Cache-Control': 'no-cache',
    'sec-ch-ua': '"Google Chrome";v="125", "Chromium";v="125",
"Not.A/Brand";v="24"',
    'sec-ch-ua-mobile': '?0',
    'sec-ch-ua-platform': '"Windows"',
    'Up tor.css('.DivTr a::attr(href)').getall()
for index in href:
    url = f'https:{index}'
    print(url)
    response = requests.get(url, headers)

    html_data = response.text
    selector = parsel.Selector(html_data)
    title = selector.css('.c_l_title h1::text').get()
```

```
    content_list = selector.css('div.noveContent p::text').getall()
    content = '\n'.join(content_list)
    with open(title + '.txt', mode='w', encoding='utf-8') as f:
        f.write(content)
```

以上案例可以获取fl小说网的免费章节，那么付费章节呢

付费章节是照片形式的存在，找到照片然后用百度云计算解析照片的文字即可，爬取付费内容是违法行为，这部分代码不能提供