

```

from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
import time
import parsel
import datetime
import csv

page_number=1

def analyze_and_save_data(data):
    global page_number
    selector = parsel.Selector(text=data)
    divs = selector.css('.sg-col-4-of-12.s-result-item.s-asin.sg-col-4-of-16.sg-col.s-widget-spacing-small.sg-col-4-of-20')
    page_position = 0 # 设置初始页面位置为 0, 并在后面设置每解析一个商品数据页面位置数加 1
    for div in divs:
        page_position += 1
        # 以下两条命令从 url 中获取页面数
        title = div.css('.s-line-clamp-1 span::text').get()
        title_parts = div.css(
            'span.a-size-base-plus.a-color-base.a-text-normal::text, span.a-size-base-plus.a-color-base.a-text-bold.a-text-normal::text').extract()
        titlecontent = ''.join(title_parts)
        price = div.css('.a-offscreen::text').get()
        original_price = div.css('.a-price.a-text-price span::text').get()
        img_url = div.css('.a-section.aok-relative.s-image-tall-aspect .s-image::attr(src)').get()
        number_of_comment11 = div.css(
            'div.a-section.a-spacing-none.a-spacing-top-micro div.a-row.a-size-small span::attr(aria-label)').extract()

        aaaa = len(number_of_comment11)
        if len(number_of_comment11) == 2:
            number_of_comment = number_of_comment11[1]
            comment = number_of_comment11[0]
        else:
            number_of_comment = '0'
            comment = '0'

        link = div.css('.a-link-normal.s-no-outline::attr(href)').get()
        prime = div.css('.a-icon.a-icon-prime.a-icon-medium::attr(aria-label)').get()

```

```

mail_service = div.css(
    '.a-row.a-size-base.a-color-secondary.s-align-children-center .a-color-base::text').get()
exclusive_to_amazon = div.css('.a-color-state.puis-light-weight-text::text').get()
small_business = div.css('.a-declarative .a-size-base.a-color-base::text').get()
sponsored = div.css('.s-label-popover-default .a-color-secondary::text').get()
coupon = div.css(
    '.a-size-base.s-highlighted-text-padding.aok-inline-block.s-coupon-highlight-color::text').get()
other_welfare = div.css('.a-badge-label-inner.a-text-ellipsis .a-badge-text::text').get()
# 获取 asin 编码
asin_code = div.css(
    '.sg-col-4-of-12.s-result-item.s-asin.sg-col-4-of-16.sg-col.s-widget-spacing-small.sg-col-4-of-20::attr(data-asin)').get()

# 设置数据爬取的具体时间
todaydate = str(datetime.date.today())
time1 = todaydate[0:4]
time2 = todaydate[5:7]
time3 = todaydate[8:]
time = time1 + '.' + time2 + '.' + time3
rounds=1
web_url=keyword
with open(f'亚马逊{todaydate}.csv', mode='a', encoding='utf-8', newline='') as f:
    csv_writer = csv.writer(f)
    csv_writer.writerow(
        [keyword, asin_code, title, titlecontent, price, original_price, img_url, comment,
number_of_comment,
        link, prime, mail_service, exclusive_to_amazon, small_business, coupon,
        other_welfare, web_url, page_position, page_number, time, sponsored, rounds])

def paqu(keyword):
    driver = webdriver.Chrome() # 创建 ChromeDriver 实例

    # 打开网页
    driver.get("https://www.amazon.com/") # 打开亚马逊官网
    wait = WebDriverWait(driver, 30) # 等待网页搜索框加载出来, 最多等待 30s
    search_input = wait.until(EC.presence_of_element_located((By.ID, 'twotabsearchtextbox')))
    # 输入检索关键词并检索
    search_input.send_keys(keyword) # 输入关键词
    submit_button = driver.find_element(By.ID, 'nav-search-submit-button') # 寻找检索 button
    submit_button.click() # 点击检索 button

```

```
# wait = WebDriverWait(driver, 10)
# wait.until(EC.presence_of_element_located((By.ID, 'search'))) # 你可以将'search'更改为包含搜索结果的适当元素
```

```
# 获取更新后的页面源代码并解析
```

```
data = driver.page_source
```

```
analyze_and_save_data(data)
```

```
time.sleep(5)
```

```
for i in range(6):
```

```
    global page_number
```

```
    page_number+=1
```

```
    next_button = driver.find_element(By.CSS_SELECTOR, '.s-pagination-next')
```

```
    next_button.click()
```

```
    time.sleep(5)
```

```
    data = driver.page_source
```

```
    analyze_and_save_data(data)
```

```
driver.quit()
```

```
for aa in keywords:
```

```
    keyword=aa
```

```
    paqu(keyword)
```

```
    print("成功爬取了关键词: "+keyword)
```