

上市公司数据采集

新浪财经新股发行：日期 / 股票名称 / 申购代码 / 申购价格。

```
In [1]: import re
import time
import requests
import pandas as pd
from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.chrome.options import Options
from selenium.webdriver.chrome.service import Service
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
```

```
In [2]: url = 'https://vip.stock.finance.sina.com.cn/corp/view/VRPD_NewStockIssue.php?page=
headers = {'User-Agent': 'Mozilla/5.0'}
```

1. 用 requests + 正则获取

```
In [3]: r = requests.get(url, headers=headers, timeout=15)
r.encoding = 'gb2312'
html = r.text

table_html = re.search(r'<table id="NewStockTable"[\s\S]*?</table>', html).group()
records = []
for row in re.findall(r'<tr[^>]*>([\s\S]*?)</tr>', table_html):
    cells = re.findall(r'<td[^>]*>([\s\S]*?)</td>', row)
    if len(cells) < 8:
        continue
    cells = [re.sub(r'<[^>]+>', '', c).replace('&nbsp;', '').strip() for c in cells]
    if not re.fullmatch(r'\d{6}', cells[1]):
        continue
    records.append({'日期': cells[3], '股票名称': cells[2], '申购代码': cells[1],
                    '申购价格': cells[7] or '--'})

df_static = pd.DataFrame(records)
df_static.head()
```

```
Out[3]:
```

	日期	股票名称	申购代码	申购价格
0	2026-05-18	长进光子	787635	0.00
1	2026-05-13	惠康科技	001237	53.26
2	2026-05-13	嘉德利	732435	15.76
3	2026-05-13	朗信电气	920220	28.29
4	2026-05-07	天海电子	001365	27.19

2. 用 selenium 获取动态数据

```
In [4]: options = Options()
options.add_argument('--headless=new')
service = Service(r'D:\mypython\myprojects\env\chromedriver-win64\chromedriver.exe')
driver = webdriver.Chrome(service=service, options=options)
```

```
In [5]: driver.get(url)
WebDriverWait(driver, 15).until(EC.presence_of_element_located((By.CSS_SELECTOR, '#
time.sleep(1)

rows = []
for tr in driver.find_elements(By.CSS_SELECTOR, '#NewStockTable tr'):
    tds = tr.find_elements(By.TAG_NAME, 'td')
    if len(tds) < 8:
        continue
    texts = [td.text.strip().replace('*', '') for td in tds]
    if not re.fullmatch(r'\d{6}', texts[1]):
        continue
    rows.append({'日期': texts[3], '股票名称': texts[2], '申购代码': texts[1],
                '申购价格': texts[7] or '--'})

df = pd.DataFrame(rows)
df.head()
```

Out[5]:

	日期	股票名称	申购代码	申购价格
--	----	------	------	------

0	2026-05-18	长进光子	787635	0.00
1	2026-05-13	朗信电气	920220	28.29
2	2026-05-13	惠康科技	001237	53.26
3	2026-05-07	天海电子	001365	27.19
4	2026-05-06	维通利	001393	30.38

```
In [6]: df.to_csv('上市公司新股申购数据.csv', index=False, encoding='utf-8-sig')
driver.quit()
print('已保存', len(df), '条')
```

已保存 40 条