## 作业

## 作业五 (选做)

目标网站: <a href="https://www.dy2018.com/html/gndy/dyz">https://www.dy2018.com/html/gndy/dyz</a>z/index.html

需求:译名、类型、评分、下载地址 (迅雷下载)

## 作业四

目标网站: <a href="http://www.piaofang.biz/">http://www.piaofang.biz/</a>

需求: 1、爬取页面所有电影名及票房,一共是100条

数据 2、保存到csv

| 4   | A             | В                |
|-----|---------------|------------------|
| - 1 | title         | 票房               |
| 2   | 阿凡达           | 2, 923, 706, 026 |
| 3   | 复仇者联盟4:终局之战   | 2, 799, 439, 100 |
| 4   | 阿凡达2:水之道      | 2, 320, 250, 281 |
| 5   | 泰坦尼克号         | 2, 264, 743, 305 |
| 6   | 星球大战:原力觉醒     | 2, 071, 310, 218 |
| 7   | 复仇者联盟3:无限战争   | 2, 052, 415, 039 |
| 8   | 蜘蛛侠:英雄无归      | 1, 921, 847, 111 |
| 9   | 侏罗纪世界         | 1, 671, 537, 444 |
| 10  | 狮子王           | 1, 663, 075, 401 |
| 11  | 复仇者联盟         | 1, 520, 538, 536 |
| 12  | 速度与激情7        | 1, 515, 341, 399 |
| 13  | 壮志凌云2:独行侠     | 1, 495, 696, 292 |
| 14  | 冰雪奇缘2         | 1, 453, 683, 476 |
| 15  | 芭比            | 1, 441, 484, 916 |
| 16  | 复仇者联盟2: 奥创纪元  | 1, 405, 018, 048 |
| 17  | 超级马力欧兄弟大电影    | 1, 361, 772, 504 |
| 18  | 黑豹            | 1, 349, 926, 083 |
| 19  | 哈利・波特与死亡圣器(下) | 1, 342, 359, 942 |
| 20  | 星球大战8:最后的绝地武士 | 1, 334, 407, 706 |
| 21  | 休罗纪世界2        | 1, 310, 466, 296 |
| 22  | 冰雪奇缘          | 1, 284, 540, 518 |
| 23  | 美女与野兽         | 1, 266, 115, 964 |
| 24  | 超人总动员2        | 1, 243, 225, 667 |
| 25  | 速度与激情8        | 1, 236, 005, 118 |
| 26  | 钢铁侠3          | 1, 215, 577, 205 |
| 27  | 小黄人大眼萌        | 1, 159, 444, 662 |
| 28  | 美国队长3:英雄内战    | 1, 155, 046, 416 |
| 29  | 海王            | 1, 148, 528, 393 |
| 30  | 指环王3:王者归来     | 1,147,633,833    |
| 31  | 007:大破天幕杀机    | 1, 142, 471, 295 |

## 作业三

目标网站: <a href="https://www.liepin.com/zhaopin/?city=41">https://www.liepin.com/zhaopin/?city=41</a>
<a href="mailto:own.com/zhaopin/?city=41">o&dq=410&pubTime=¤tPage=0&pageSize=40&key=python&suggestTag=&workYearCode=&compId=&compName=&compTag=&industry=&salary=&jobKind=&compScale=&compKind=&compStage=&eduLevel=&otherCity=&sfrom=search\_job\_pc&ckld=hwmufbahh4eacxrzevqwbxwfh36ui9d9&scene=input&skld=hwmufbahh4eacxrzevqwbxwfh36ui9d9&scene=input&skld=hwmufbahh4eacxrzevqwbxwfh36ui9d9&suggestId=</a>

需求: 翻页爬取前3页的职位名称、薪资、公司名称

#### 1. 先获取第一页数据

目标url: <a href="https://api-c.liepin.com/api/com.liepin.s">https://api-c.liepin.com/api/com.liepin.s</a> earchfront4c.pc-search-job

响应数据类型: json

请求方式: post请求

#### • 效果图

## 作业二

目标网站: <a href="https://vip.stock.finance.sina.com.cn/mkt/#bjs\_root">https://vip.stock.finance.sina.com.cn/mkt/#bjs\_root</a> (这个url不是真正抓取数据的url)需求: 打印输出股票名称,最新价,跌涨额,涨跌幅,买入,卖出,昨收,今开等数据

• 效果图

```
特锐德 19.720 0.28 1.44%
神州泰岳 10.430 0.16 1.558%
乐普医疗 19.890 -0.19 -0.946%
南风股份 5.670 0 0%
探路者 7.780 0.04 0.517%
莱美药业 3.590 -0.05 -1.374%
汉威科技 17.060 0.46 2.771%
天海防务 4.940 0.01 0.203%
安科生物 10.290 0.01 0.097%
*ST豆神 2.330 0.01 0.431%
鼎汉技术 6.940 0.02 0.289%
```

### 作业一

# 由于百度容易百度验证出现问题,给大家换了一个题目哈

目标网站:

https://www.sogou.com/web?query=java

https://www.sogou.com/web?query=python

需求:用户搜索不同内容,保存不同的网页源码。

比如搜索java, 保存java源码

搜索python, 保存python源码

1. 当前的网页源码保存下来

确定url: