例：程序crawler\_test/crawler\_test2.py

# 爬取斗破苍穹小说  
import requests  
import re  
import time  
  
headers = {'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; WOW64) '  
 'AppleWebKit/537.36 (KHTML, like Gecko) '  
 'Chrome/65.0.3325.181 Safari/537.36'}  
  
file\_object = open('doupo.txt', 'a+')  
  
  
def get\_info(url):  
 res = requests.get(url, headers=headers)  
 if res.status\_code == 200: # 请求成功  
 contents = re.findall('<p>(.\*?)</p>', res.content.decode('utf-8'))  
 for content in contents:  
 file\_object.write(content + '\n')  
 else:  
 pass  
  
  
if \_\_name\_\_ == '\_\_main\_\_':  
 urls = ['http://www.doupoxs.com/doupocangqiong/{0}.html'.format(str(i)) for i in range(1, 100)]  
 for url in urls:  
 get\_info(url)  
 time.sleep(1)  
  
 file\_object.close()

例：程序crawler\_test/crawler\_test3.py

# 爬取糗事百科  
import requests  
import re  
import time  
import crawler\_test1  
  
info\_lists = []  
  
  
def judge\_sex(sex):  
 if sex == 'womenIcon':  
 return '女'  
 else:  
 return '男'  
  
  
def get\_info(url):  
 res = requests.get(url, headers=crawler\_test1.headers)  
 # 用户ID  
 ids = re.findall('<h2>(.\*?)</h2>', res.text, re.S) # 换行匹配  
 # 用户等级  
 levels = re.findall('<div class="articleGender \D+Icon">(.\*?)</div>', res.text, re.S)  
 # 性别  
 sexs = re.findall('<div class="articleGender (.\*?)">', res.text, re.S)  
 # 内容  
 contents = re.findall('<div class="content">.\*?<span>(.\*?)</span>', res.text, re.S)  
 # 好笑  
 laughs = re.findall('<span class="stats-vote"><i class="number">(\d+)</i>', res.text, re.S)  
 # 评论  
 comments = re.findall('<i class="number">(\d+)</i> 评论', res.text, re.S)  
  
 for id, level, sex, content, laugh, comment in zip(ids, levels, sexs, contents, laughs, comments):  
 info = {'id': id, 'level': level, 'sex': judge\_sex(sex), 'content': content,  
 'laugh': laugh, 'comment': comment}  
 info\_lists.append(info)  
  
  
if \_\_name\_\_ == '\_\_main\_\_':  
 url = 'https://www.qiushibaike.com/text/page/12/'  
 get\_info(url)  
 print(info\_lists)