```python
import requests
import time
from lxml import etree


url = 'https://book.douban.com/top250'
data = requests.get(url).text
s = etree.HTML(data)

file = s.xpath('//*[@id="content"]/div/div[1]/div/table')

for div in file:
    title = div.xpath("./tr/td[2]/div[1]/a/@title")[0]
    score = div.xpath("./tr/td[2]/div[2]/span[2]/text()")[0]
    num = div.xpath("./tr/td[2]/div[2]/span[3]/text()")[0].strip("(").strip().strip(")
    scribe = div.xpath("./tr/td[2]/p[2]/span/text()")

    time.sleep(2)

    print("{}{}{}{}".format(title,  score, num, scribe[0]))
```

for div in file

Run: 静态

为你，千千万万遍
解忧杂货店8.5420885人评价
一碗精心熬制的东野牌鸡汤，拒绝很难
小王子9.0356745人评价
献给长成了大人的孩子们
白夜行9.1311227人评价
暗夜独行的残破灵魂，爱与恶本就难分难舍
围城8.9261298人评价
对于"人艰不拆"四个字最彻底的违抗
三体8.8257336人评价
你我不过都是虫子

for div in file

静态

E:\网站\实验\Scripts\python.exe E:/高凯作业/爬虫/静态.py
追风筝的人8.9434316人评价
为你，千千万万遍
解忧杂货店8.5420885人评价
一碗精心熬制的东野牌鸡汤，拒绝很难
小王子9.0356745人评价
献给长成了大人的孩子们
白夜行9.1311227人评价

```python
import requests

import time

from lxml import etree


url = 'https://book.douban.com/top250'

data = requests.get(url).text

s = etree.HTML(data)


file = s.xpath('//*[@id="content"]/div/div[1]/div/table')


for div in file:
    title = div.xpath("./tr/td[2]/div[1]/a/@title")[0]
    score = div.xpath("./tr/td[2]/div[2]/span[2]/text()")[0]
    num =
div.xpath("./tr/td[2]/div[2]/span[3]/text()")[0].strip("(").strip().strip(")")
```

```python
scribe = div.xpath("./tr/td[2]/p[2]/span/text()")

time.sleep(2)

print("{}{}{}{}".format(title, score, num, scribe[0]))
```

```python
import time
from selenium import webdriver
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import WebDriverWait

from lxml import etree

browser = webdriver.Chrome()
browser.get("https://www.baidu.com")

wait = WebDriverWait(browser, 50)


def search():
    browser.get('https://www.jd.com/')
    try:
        input = wait.until(
            EC.presence_of_all_elements_located((By.CSS_SELECTOR, "#key"))
        ) # llist
        submit = wait.until(
            EC.element_to_be_clickable((By.CSS_SELECTOR, "#search > div > div.form > button
```

Run: 动态

```
img: //img10.360buyimg.com/n1/s200x200_jfs/t23722/106/2427071620/128371/2d7b202a/5b7e6e0bNc42f7eea.jpg
title: Python从菜鸟到高手
price: 115.20
commit 14万+
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
img: //img12.360buyimg.com/n1/s200x200_jfs/t11800/244/1971230731/124439/a9f98805/5a0db741N8c4a2e72.jpg
```

Run: 动态

```
E:\网站\实验\Scripts\python.exe E:/高凯作业/爬虫/动态.py
第 1 页:
img: //img14.360buyimg.com/n1/s200x200_jfs/t1/15399/24/5595/317531/5c3fe258E00795253/de690946fa5d
title: Python从入门到项目实践（全彩版）
price: 70.60
commit 9万+
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
img: //img14.360buyimg.com/n1/s200x200_jfs/t17953/201/1450663539/451183/3262b8de/5acb3627N8191c867
title: Python编程 从入门到实践
price: 61.40
commit 21万+
```

```python
import time
from selenium import webdriver
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import WebDriverWait

from lxml import etree

browser = webdriver.Chrome()
browser.get("https://www.baidu.com")

wait = WebDriverWait(browser, 50)


def search():
    browser.get('https://www.jd.com/')
    try:
        input = wait.until(
            EC.presence_of_all_elements_located((By.CSS_SELECTOR, "#key"))
        ) # llist
        submit = wait.until(
```

```python
            EC.element_to_be_clickable((By.CSS_SELECTOR, "#search > div > div.form >
button"))
        )
        # input = browser.find_element_by_id('key')
        input[0].send_keys('python')
        submit.click()

        total = wait.until(
            EC.presence_of_all_elements_located(
                (By.CSS_SELECTOR, '#J_bottomPage > span.p-skip > em:nth-child(1) >
b')
            )
        )
        html = browser.page_source
        prase_html(html)
        return total[0].text
    except TimeoutError:
        search()


def next_page(page_number):
    try:
        # 滑动到底部，加载出后三十个货物信息
        browser.execute_script("window.scrollTo(0, document.body.scrollHeight);")
        time.sleep(10)
        # 翻页动作
        button = wait.until(
            EC.element_to_be_clickable((By.CSS_SELECTOR, '#J_bottomPage >
span.p-num > a.pn-next > em'))
        )
        button.click()
        wait.until(
            EC.presence_of_all_elements_located((By.CSS_SELECTOR, "#J_goodsList >
ul > li:nth-child(60)"))
        )
        # 判断翻页成功
        wait.until(
            EC.text_to_be_present_in_element((By.CSS_SELECTOR, "#J_bottomPage >
span.p-num > a.curr"), str(page_number))
        )
        html = browser.page_source
        prase_html(html)
    except TimeoutError:
        return next_page(page_number)
```

```python
def prase_html(html):
    html = etree.HTML(html)
    items = html.xpath('//li[@class="gl-item"]')
    for i in range(len(items)):
        if html.xpath('//div[@class="p-img"]//img')[i].get('data-lazy-img') !=
"done":
            print("img:",
html.xpath('//div[@class="p-img"]//img')[i].get('data-lazy-img'))
        else:
            print("img:", html.xpath('//div[@class="p-img"]//img')[i].get('src'))
        print("title:",
html.xpath('//div[@class="p-name"]//em')[i].xpath('string(.)'))
        print("price:", html.xpath('//div[@class="p-price"]//i')[i].text)
        print("commit", html.xpath('//div[@class="p-commit"]//a')[i].text)

    print("++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++")


def main():
    print("第", 1, "页：")
    total = int(search())
    for i in range(2, total + 1):
        time.sleep(3)
        print("第", i, "页：")
        next_page(i)


if __name__ == "__main__":
    main()
```

```python
from lxml import etree
import requests

def gk():
    url = 'https://book.douban.com/top250'
    data = requests.get(url).text
    s=etree.HTML(data)



file=s.xpath('//*[@id="content"]/div/div[1]/div/table/tr/td[2]/div[1]/a/@title')
    print(file)
    return file
```
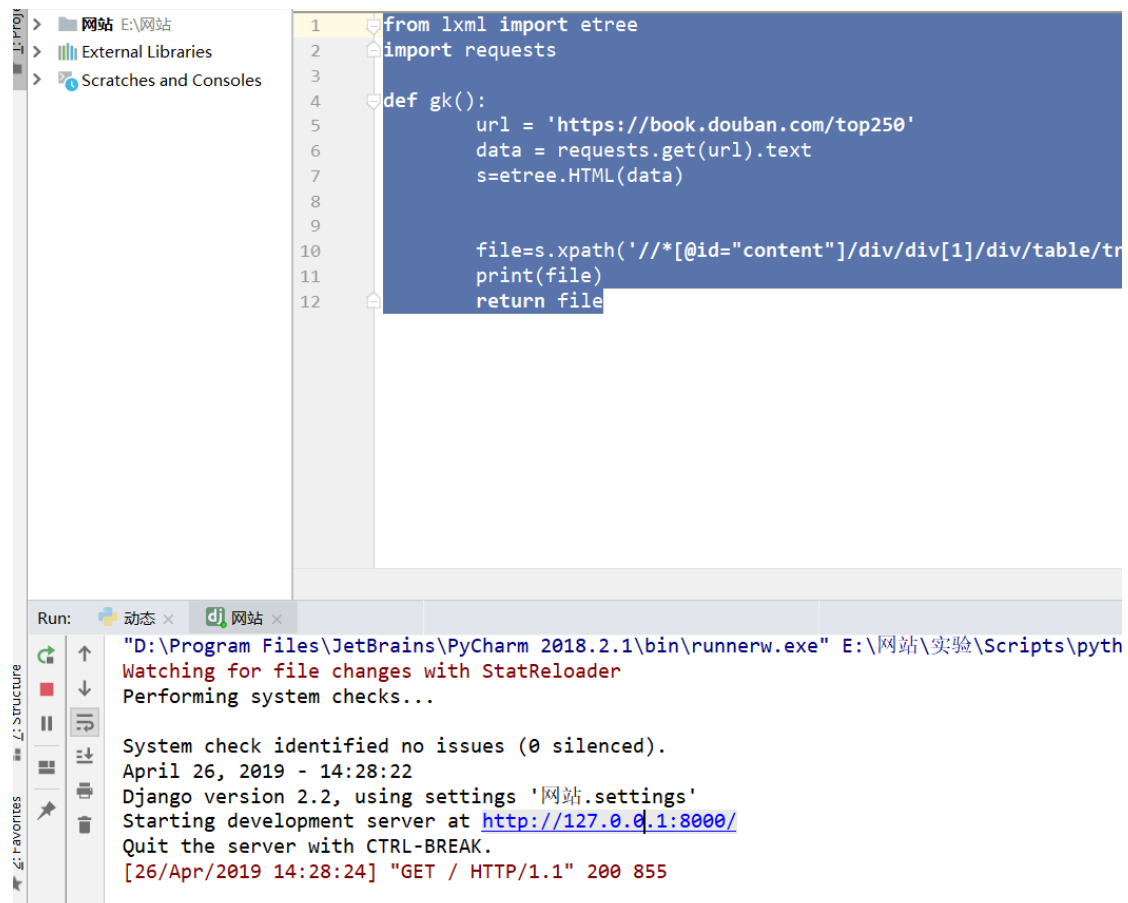
追风筝的人

解忧杂货店

小王子

白夜行

围城

三体

嫌疑人X的献身

活着

挪威的森林

百年孤独

红楼梦

```python
from lxml import etree
import requests

def gk():
        url = 'https://book.douban.com/top250'
        data = requests.get(url).text
        s=etree.HTML(data)


        file=s.xpath('//*[@id="content"]/div/div[1]/div/table/tr
        print(file)
        return file
```

```
Run:    动态 ×    网站 ×

"D:\Program Files\JetBrains\PyCharm 2018.2.1\bin\runnerw.exe" E:\网站\实验\Scripts\pyth
Watching for file changes with StatReloader
Performing system checks...

System check identified no issues (0 silenced).
April 26, 2019 - 14:28:22
Django version 2.2, using settings '网站.settings'
Starting development server at http://127.0.0.1:8000/
Quit the server with CTRL-BREAK.
[26/Apr/2019 14:28:24] "GET / HTTP/1.1" 200 855
```

期末作业：12306 自动登录

```python
from selenium import webdriver
import time
import requests
from hashlib import md5
import re
import base64
from selenium.webdriver.common.action_chains import ActionChains


class Chaojiying_Client(object):

    def __init__(self, username, password, soft_id):
        self.username = username
        password = password.encode('utf8')
        self.password = md5(password).hexdigest()
        self.soft_id = soft_id
        self.base_params = {
            'user': self.username,
            'pass2': self.password,
```

```python
            'softid': self.soft_id,
        }
        self.headers = {
            'Connection': 'Keep-Alive',
            'User-Agent': 'Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 5.1;
Trident/4.0)',
        }

    def PostPic(self, im, codetype):
        params = {
            'codetype': codetype,
        }
        params.update(self.base_params)
        files = {'userfile': ('12306.jpg', im)}
        r =
requests.post('http://upload.chaojiying.net/Upload/Processing.php',
data=params, files=files, headers=self.headers)
        return r.json()

    def ReportError(self, im_id):
        params = {
            'id': im_id,
        }
        params.update(self.base_params)
        r =
requests.post('http://upload.chaojiying.net/Upload/ReportError.php',
data=params, headers=self.headers)
        return r.json()


class Login:
    def __init__(self, url, username, password):
        self.url = url
        self.username = username
        self.password = password

    def login(self):
        self.browser = webdriver.Chrome()
        self.browser.get(self.url)
        time.sleep(2)
        login_select =
self.browser.find_element_by_class_name('login-hd-account')
        login_select.click()
        user = self.browser.find_element_by_id('J-userName')
        word = self.browser.find_element_by_id('J-password')
```

```python
            user.send_keys(self.username)
            word.send_keys(self.password)

    def get_pic(self):
        tag = self.browser.find_element_by_class_name('imgCode')
        temp = tag.get_attribute('src')
        b64_pic = re.sub(r'data:image/jpg;base64,', '', temp)
        pic = base64.b64decode(b64_pic)
        return pic

    def click(self, j):
        temp = j.get('pic_str')
        locations = [list(map(int, i.split(','))) for i in temp.split('|')]
        for location in locations:

ActionChains(self.browser).move_to_element_with_offset(self.browser.find_element_by_class_name('imgCode'),
                                                       location[0],
location[1]).click().perform()
            time.sleep(5)
        self.browser.find_element_by_id('J-login').click()

    def get_cookies(self):
        return self.browser.get_cookies()


if __name__ == '__main__':
    a = Login('https://kyfw.12306.cn/otn/resources/login.html',
'15558977180', 'gaokai5221')
    a.login()
    time.sleep(2)
    chaojiying = Chaojiying_Client('15858772795', 'chenxiangniu110.',
'900049')
    im = a.get_pic()
    print(im)
    z = chaojiying.PostPic(im, 9004)
    print(z)
    a.click(z)
    print(a.get_cookies())
```