import requests  
from bs4 import BeautifulSoup  
import re  
import os  
  
#获取所有页面HTML  
def get\_pages(url):  
 contents = []  
 start = 0  
 for i in range(10):  
 response = requests.get(url = url,params = {"start":start,"filter":""})  
 html = response.text  
 bs = BeautifulSoup(html, "html.parser")  
 content = bs.find('ol', class\_ = "grid\_view")  
 start += 25  
 contents.append(content)  
 return contents  
  
#提取电影信息  
def extra\_info(content):  
 info\_list = []  
 tag\_list = content.find\_all('li')  
 for i in tag\_list:  
 name = i.find('span',class\_ = 'title').string  
 star = i.find('span', class\_ = 'rating\_num').string  
 comment = i.find('span',property="v:best").next\_sibling.next\_sibling.string  
 comment = re.search("(.\*)人评价",comment).group(1)  
 # print(comment)  
 a = name,star,comment  
 info\_list.append(a)  
 return info\_list  
  
#保存到文件  
def dump\_file(info\_list):  
 file\_name = "top\_250.txt"  
 fp = open(file\_name,'a+')  
 for film\_item in info\_list:  
 for item in film\_item:  
 fp.write(item+'\n')  
 fp.write('------'\*5+'\n')  
 fp.close()  
  
if \_\_name\_\_ == "\_\_main\_\_":  
 file\_name = "top\_250.txt"  
 contents = get\_pages("https://movie.douban.com/top250")  
 #如果文件存在，则删除  
 if os.path.isfile(file\_name):  
 os.remove(file\_name)  
 for content in contents:  
 info\_list = extra\_info(content)  
 dump\_file(info\_list)

该实例只是简单的静态页面爬取