import pprint

import time

from lxml import etree

import csv

import requests

with open('马蜂窝.csv', 'w', newline='', encoding='utf-8-sig') as csvfile:

fieldnames = ['标题', '发布时间', '正文']

writer = csv.DictWriter(csvfile, fieldnames=fieldnames)

writer.writeheader()

# 遍历每一页

for i in range(1, 11):

cookies = {

'mfw\_uuid': '67b828ac-8078-040c-167d-29998bc29188',

'oad\_n': 'a%3A3%3A%7Bs%3A3%3A%22oid%22%3Bi%3A1029%3Bs%3A2%3A%22dm%22%3Bs%3A15%3A%22www.mafengwo.cn%22%3Bs%3A2%3A%22ft%22%3Bs%3A19%3A%222025-02-21+15%3A18%3A04%22%3B%7D',

'\_\_mfwa': '1740122285874.58494.1.1740122285874.1740122285874',

'\_\_mfwlv': '1740122285',

'\_\_mfwvn': '1',

'uva': 's%3A130%3A%22a%3A3%3A%7Bs%3A2%3A%22lt%22%3Bi%3A1740122285%3Bs%3A10%3A%22last\_refer%22%3Bs%3A62%3A%22https%3A%2F%2Fwww.mafengwo.cn%2Ftravel-scenic-spot%2Fmafengwo%2F12931.html%22%3Bs%3A5%3A%22rhost%22%3BN%3B%7D%22%3B',

'\_\_mfwurd': 'a%3A3%3A%7Bs%3A6%3A%22f\_time%22%3Bi%3A1740122285%3Bs%3A9%3A%22f\_rdomain%22%3Bs%3A15%3A%22www.mafengwo.cn%22%3Bs%3A6%3A%22f\_host%22%3Bs%3A3%3A%22www%22%3B%7D',

'\_\_mfwuuid': '67b828ac-8078-040c-167d-29998bc29188',

'bottom\_ad\_status': '0',

'PHPSESSID': 'trdk69mpe4prr60e8vei05ojd4',

'Hm\_lvt\_8288b2ed37e5bc9b4c9f7008798d2de0': '1740122286,1740123813',

'HMACCOUNT': 'D1D7F8ADA1F46996',

'\_\_mfwothchid': 'referrer%7Copen.weixin.qq.com',

'\_\_omc\_chl': '',

'\_\_mfwc': 'referrer%7Copen.weixin.qq.com',

'login': 'mafengwo',

'mafengwo': 'ca954a4754e88a50cd35d50921b19d76\_71493921\_67b831e59b47e1.95581597\_67b831e59b4883.38857285',

'x-waf-captcha-referer': 'https%3A%2F%2Fwww.mafengwo.cn%2F',

'uol\_throttle': '71493921',

'mfw\_uid': '71493921',

'\_\_omc\_r': '',

'\_\_mfwb': 'cd22471cc6ad.14.direct',

'\_\_mfwlt': '1740124674',

'Hm\_lpvt\_8288b2ed37e5bc9b4c9f7008798d2de0': '1740124675',

'w\_tsfp': 'ltvuV0MF2utBvS0Q7arolk+oHz0vcj04h0wpEaR0f5thQLErU5mB0Yd4us3+Nnbf5Mxnvd7DsZoyJTLYCJI3dwMWR8TFJYEQiV+ZxoRx3NtHBxdvEMiNXFQWcegn7mNBKXhCNxS00jA8eIUd379yilkMsyN1zap3TO14fstJ019E6KDQmI5uDW3HlFWQRzaLbjcMcuqPr6g18L5a5W2J4gn6LVp9A7lCgxaS1C5NWHl36Ra7fOhZYEr7Js/8SqA=',

}

headers = {

'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/112.0.0.0 Safari/537.36 QuarkPC/2.0.7.222',

'Accept': 'application/json, text/javascript, \*/\*; q=0.01',

'sec-ch-ua': '"Not:A-Brand";v="99", "Chromium";v="112"',

'x-requested-with': 'XMLHttpRequest',

'sec-ch-ua-mobile': '?0',

'sec-ch-ua-platform': '"Windows"',

'origin': 'https://www.mafengwo.cn',

'sec-fetch-site': 'same-origin',

'sec-fetch-mode': 'cors',

'sec-fetch-dest': 'empty',

'accept-language': 'zh-CN,zh;q=0.9',

'referer': 'https://www.mafengwo.cn/travel-scenic-spot/mafengwo/12931.html',

# Requests sorts cookies= alphabetically

# 'Cookie': 'mfw\_uuid=67b828ac-8078-040c-167d-29998bc29188; oad\_n=a%3A3%3A%7Bs%3A3%3A%22oid%22%3Bi%3A1029%3Bs%3A2%3A%22dm%22%3Bs%3A15%3A%22www.mafengwo.cn%22%3Bs%3A2%3A%22ft%22%3Bs%3A19%3A%222025-02-21+15%3A18%3A04%22%3B%7D; \_\_mfwa=1740122285874.58494.1.1740122285874.1740122285874; \_\_mfwlv=1740122285; \_\_mfwvn=1; uva=s%3A130%3A%22a%3A3%3A%7Bs%3A2%3A%22lt%22%3Bi%3A1740122285%3Bs%3A10%3A%22last\_refer%22%3Bs%3A62%3A%22https%3A%2F%2Fwww.mafengwo.cn%2Ftravel-scenic-spot%2Fmafengwo%2F12931.html%22%3Bs%3A5%3A%22rhost%22%3BN%3B%7D%22%3B; \_\_mfwurd=a%3A3%3A%7Bs%3A6%3A%22f\_time%22%3Bi%3A1740122285%3Bs%3A9%3A%22f\_rdomain%22%3Bs%3A15%3A%22www.mafengwo.cn%22%3Bs%3A6%3A%22f\_host%22%3Bs%3A3%3A%22www%22%3B%7D; \_\_mfwuuid=67b828ac-8078-040c-167d-29998bc29188; bottom\_ad\_status=0; PHPSESSID=trdk69mpe4prr60e8vei05ojd4; Hm\_lvt\_8288b2ed37e5bc9b4c9f7008798d2de0=1740122286,1740123813; HMACCOUNT=D1D7F8ADA1F46996; \_\_mfwothchid=referrer%7Copen.weixin.qq.com; \_\_omc\_chl=; \_\_mfwc=referrer%7Copen.weixin.qq.com; login=mafengwo; mafengwo=ca954a4754e88a50cd35d50921b19d76\_71493921\_67b831e59b47e1.95581597\_67b831e59b4883.38857285; x-waf-captcha-referer=https%3A%2F%2Fwww.mafengwo.cn%2F; uol\_throttle=71493921; mfw\_uid=71493921; \_\_omc\_r=; \_\_mfwb=cd22471cc6ad.14.direct; \_\_mfwlt=1740124674; Hm\_lpvt\_8288b2ed37e5bc9b4c9f7008798d2de0=1740124675; w\_tsfp=ltvuV0MF2utBvS0Q7arolk+oHz0vcj04h0wpEaR0f5thQLErU5mB0Yd4us3+Nnbf5Mxnvd7DsZoyJTLYCJI3dwMWR8TFJYEQiV+ZxoRx3NtHBxdvEMiNXFQWcegn7mNBKXhCNxS00jA8eIUd379yilkMsyN1zap3TO14fstJ019E6KDQmI5uDW3HlFWQRzaLbjcMcuqPr6g18L5a5W2J4gn6LVp9A7lCgxaS1C5NWHl36Ra7fOhZYEr7Js/8SqA=',

}

params = {

'act': 'get\_travellist',

}

data = {

'mddid': '12931',

'pageid': 'mdd\_index',

'sort': '1',

'cost': '0',

'days': '0',

'month': '0',

'tagid': '0',

'page': i,

'\_ts': int(time.time()\*1000),

'\_sn': '2bcfaf71b6',

}

response = requests.post('https://www.mafengwo.cn/gonglve/ajax.php', params=params, cookies=cookies, headers=headers, data=data,verify=False)

pprint.pprint(response.json()['list'])

tree = etree.HTML(response.json()['list'])

items = tree.xpath('//div[@class="tn-item clearfix"]')

for item in items:

link = item.xpath('.//a[@class="title-link"]/@href')[0]

title = item.xpath('.//a[@class="title-link"]/text()')[0].strip()

print(title)

# views = item.xpath('.//span[@class="tn-nums"]/text()[1]')[0].strip()

# time = item.xpath('.//span[contains(@class, "time")]/text()')

link = 'https://www.mafengwo.cn'+link

cookies = {

'HMACCOUNT': '865B3F58A82B3078',

'Hm\_lpvt\_8288b2ed37e5bc9b4c9f7008798d2de0': '1740122951',

'Hm\_lvt\_8288b2ed37e5bc9b4c9f7008798d2de0': '1740122490',

'PHPSESSID': 'ntmsg9tvn8c0smr00hanpcg5u5',

'\_\_mfwa': '1740122489766.56087.1.1740122489766.1740122489766',

'\_\_mfwb': '4935ad5c6f53.4.direct',

'\_\_mfwc': 'direct',

'\_\_mfwlt': '1740122950',

'\_\_mfwlv': '1740122489',

'\_\_mfwurd': 'a%3A3%3A%7Bs%3A6%3A%22f\_time%22%3Bi%3A1740122489%3Bs%3A9%3A%22f\_rdomain%22%3Bs%3A15%3A%22www.mafengwo.cn%22%3Bs%3A6%3A%22f\_host%22%3Bs%3A3%3A%22www%22%3B%7D',

'\_\_mfwuuid': '67b82978-3517-8111-040d-1a70122e39ce',

'\_\_mfwvn': '1',

'bottom\_ad\_status': '0',

'mfw\_uuid': '67b82978-3517-8111-040d-1a70122e39ce',

'oad\_n': 'a%3A3%3A%7Bs%3A3%3A%22oid%22%3Bi%3A1029%3Bs%3A2%3A%22dm%22%3Bs%3A15%3A%22www.mafengwo.cn%22%3Bs%3A2%3A%22ft%22%3Bs%3A19%3A%222025-02-21+15%3A21%3A28%22%3B%7D',

'uva': 's%3A130%3A%22a%3A3%3A%7Bs%3A2%3A%22lt%22%3Bi%3A1740122489%3Bs%3A10%3A%22last\_refer%22%3Bs%3A62%3A%22https%3A%2F%2Fwww.mafengwo.cn%2Ftravel-scenic-spot%2Fmafengwo%2F12931.html%22%3Bs%3A5%3A%22rhost%22%3BN%3B%7D%22%3B',

'w\_tsfp': 'ltvuV0MF2utBvS0Q7arolkmnEToucT84h0wpEaR0f5thQLErU5mB0Yd4vMLwM3Xe5sxnvd7DsZoyJTLYCJI3dwNHFpqSJokV2guXloUm3dhFVEUxFZnYXQRNI+995WJHKnhCNxS00jA8eIUd379yilkMsyN1zap3TO14fstJ019E6KDQmI5uDW3HlFWQRzaLbjcMcuqPr6g18L5a5W3csFLzLV0nVr5GgxSS1yweXHtysBO6J+BdM0+scZynSqA=',

}

headers = {

'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/121.0.0.0 Safari/537.36',

'Accept': 'application/json, text/javascript, \*/\*; q=0.01',

# 'Accept-Encoding': 'gzip, deflate',

'sec-ch-ua': '"Not A(Brand";v="99", "Google Chrome";v="121", "Chromium";v="121"',

'x-requested-with': 'XMLHttpRequest',

'sec-ch-ua-mobile': '?0',

'sec-ch-ua-platform': '"Windows"',

'origin': 'https://www.mafengwo.cn',

'sec-fetch-site': 'same-origin',

'sec-fetch-mode': 'cors',

'sec-fetch-dest': 'empty',

'referer': 'https://www.mafengwo.cn/travel-scenic-spot/mafengwo/12931.html',

'accept-language': 'zh-CN,zh;q=0.9',

# Requests sorts cookies= alphabetically

# 'Cookie': 'HMACCOUNT=865B3F58A82B3078; Hm\_lpvt\_8288b2ed37e5bc9b4c9f7008798d2de0=1740122951; Hm\_lvt\_8288b2ed37e5bc9b4c9f7008798d2de0=1740122490; PHPSESSID=ntmsg9tvn8c0smr00hanpcg5u5; \_\_mfwa=1740122489766.56087.1.1740122489766.1740122489766; \_\_mfwb=4935ad5c6f53.4.direct; \_\_mfwc=direct; \_\_mfwlt=1740122950; \_\_mfwlv=1740122489; \_\_mfwurd=a%3A3%3A%7Bs%3A6%3A%22f\_time%22%3Bi%3A1740122489%3Bs%3A9%3A%22f\_rdomain%22%3Bs%3A15%3A%22www.mafengwo.cn%22%3Bs%3A6%3A%22f\_host%22%3Bs%3A3%3A%22www%22%3B%7D; \_\_mfwuuid=67b82978-3517-8111-040d-1a70122e39ce; \_\_mfwvn=1; bottom\_ad\_status=0; mfw\_uuid=67b82978-3517-8111-040d-1a70122e39ce; oad\_n=a%3A3%3A%7Bs%3A3%3A%22oid%22%3Bi%3A1029%3Bs%3A2%3A%22dm%22%3Bs%3A15%3A%22www.mafengwo.cn%22%3Bs%3A2%3A%22ft%22%3Bs%3A19%3A%222025-02-21+15%3A21%3A28%22%3B%7D; uva=s%3A130%3A%22a%3A3%3A%7Bs%3A2%3A%22lt%22%3Bi%3A1740122489%3Bs%3A10%3A%22last\_refer%22%3Bs%3A62%3A%22https%3A%2F%2Fwww.mafengwo.cn%2Ftravel-scenic-spot%2Fmafengwo%2F12931.html%22%3Bs%3A5%3A%22rhost%22%3BN%3B%7D%22%3B; w\_tsfp=ltvuV0MF2utBvS0Q7arolkmnEToucT84h0wpEaR0f5thQLErU5mB0Yd4vMLwM3Xe5sxnvd7DsZoyJTLYCJI3dwNHFpqSJokV2guXloUm3dhFVEUxFZnYXQRNI+995WJHKnhCNxS00jA8eIUd379yilkMsyN1zap3TO14fstJ019E6KDQmI5uDW3HlFWQRzaLbjcMcuqPr6g18L5a5W3csFLzLV0nVr5GgxSS1yweXHtysBO6J+BdM0+scZynSqA=',

}

res = requests.get(link,cookies=cookies,headers=headers,verify=False)

# print(res.text)

tree = etree.HTML(res.text)

times = tree.xpath('//li[@class="time"]/text()[2]')

ptext = tree.xpath('//div[@class="content"]//p/text()')

text = ''.join(ptext).replace('\xa0', '').replace('\n', '').replace('\r', '').strip()

# 写入CSV

dit = {'标题': title, '发布时间': times, '正文': text}

print(dit)

writer.writerow(dit)

import time

import requests

from lxml import etree

import csv

cookies = {

'QN205': 'organic',

'QN277': 'organic',

'QN48': 'tc\_864b73104c95f701\_195273d7c65\_8d4a',

'QN269': 'E54C6954F01E11EF8B5E36FFA2C56A45',

'QN1': '00013a8034fc6b76e19864ce',

'qunar-assist': '{%22version%22:%2220211215173359.925%22%2C%22show%22:false%2C%22audio%22:false%2C%22speed%22:%22middle%22%2C%22zomm%22:1%2C%22cursor%22:false%2C%22pointer%22:false%2C%22bigtext%22:false%2C%22overead%22:false%2C%22readscreen%22:false%2C%22theme%22:%22default%22}',

'csrfToken': 'HFuwyHRmCmQpVrIQ2ZDlfqovAVK1zMiB',

'QN300': 'common.qunarzz.com',

'\_i': 'DFiEZnT0sZvw1b\_e\_fiAVRLHoLyw',

'\_vi': '2eFz\_wcXmOlaxT\_QMqkSoRq\_y7u40w977IHi1E6KZyPgvWQ3N59F91m7ZUYvdV-NBcozxM\_\_jMiklkzxU-NicslKMtP8Cube5XeSQaWjrrSojBc5nEzt9AllgWTb8zXnzOQEXyDiBkSMi7yq9JwWaGmP0DwdIg0IN-2JPZcF3XU8',

'ariaDefaultTheme': 'null',

'Hm\_lvt\_c56a2b5278263aa647778d304009eafc': '1740120096,1740133950',

'HMACCOUNT': '8179753F80C7F3D3',

'fid': 'f78d4e03-8c9c-4f56-9fc1-7a4d95cf981e',

'viewdist': '300079-1',

'uld': '1-300079-1-1740133953',

'viewbook': '6953157|7748813|7748813|7748813|7487190',

'JSESSIONID': 'CBA84FA384E2D16CF67F910898D7CD94',

'QN271': '65c9535d-fd9c-4bdb-872b-7a26834b2b76',

'QN267': '12710563478992fd83',

'Hm\_lpvt\_c56a2b5278263aa647778d304009eafc': '1740133968',

}

headers = {

'accept': 'text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,image/apng,\*/\*;q=0.8,application/signed-exchange;v=b3;q=0.7',

'accept-language': 'zh-CN,zh;q=0.9,en;q=0.8,en-GB;q=0.7,en-US;q=0.6',

'cache-control': 'max-age=0',

'priority': 'u=0, i',

'referer': 'https://travel.qunar.com/',

'sec-ch-ua': '"Not(A:Brand";v="99", "Microsoft Edge";v="133", "Chromium";v="133"',

'sec-ch-ua-mobile': '?0',

'sec-ch-ua-platform': '"Windows"',

'sec-fetch-dest': 'document',

'sec-fetch-mode': 'navigate',

'sec-fetch-site': 'same-origin',

'sec-fetch-user': '?1',

'upgrade-insecure-requests': '1',

'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/133.0.0.0 Safari/537.36 Edg/133.0.0.0',

}

with open('去哪.csv', 'w', newline='', encoding='utf-8-sig') as csvfile:

fieldnames = ['标题', '发布时间', '正文']

writer = csv.DictWriter(csvfile, fieldnames=fieldnames)

writer.writeheader()

# 遍历每一页

for i in range(1, 50):

url = f'https://travel.qunar.com/travelbook/list/23-ningxia-298293/hot\_heat/{i}.htm'

try:

response = requests.get(url, cookies=cookies, headers=headers)

time.sleep(2)

tree = etree.HTML(response.text)

print(f"正在处理：{url}")

# 获取每个文章的链接

li\_list = tree.xpath('//ul[@class="b\_strategy\_list "]/li')

for li in li\_list:

try:

title = li.xpath('./h2/a/text()')[0].strip() # 获取标题

ahref = li.xpath('./h2/a/@href')[0].split('/')[2]

ahref = 'https://travel.qunar.com/travelbook/note/' + ahref

print(f"找到文章：{title} - {ahref}")

# 请求文章详情页

res = requests.get(ahref, cookies=cookies, headers=headers)

tree = etree.HTML(res.text)

try:

# 提取发布时间

times = tree.xpath('//\*[@id="js\_mainleft"]/div[2]/ul/li[1]/p/span[2]/text()')[0].strip()

except:

times = '无日期'

try:

# 提取正文内容

text = tree.xpath('/html/body//div[@class="b\_panel\_schedule"]//p//text()')

text = ''.join(text).replace('\xa0', '').replace('\n', '').replace('\r', '').strip()

except:

text = '无内容'

# 保存数据

dit = {'标题': title, '发布时间': times, '正文': text}

print(dit)

writer.writerow(dit)

except Exception as e:

print(f"处理文章时出错：{e}")

continue # 跳过当前文章，继续处理下一个

except Exception as e:

print(f"处理第 {i} 页时出错：{e}")

continue # 跳过当前页面，继续处理下一页

print("抓取完成！")

ï»¿

import requests

from lxml import etree

import csv

# å®šä¹‰è¯·æ±‚å¤´å’ŒCookie

headers = {

'authority': 'you.ctrip.com',

'accept': 'text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,image/apng,\*/\*;q=0.8,application/signed-exchange;v=b3;q=0.7',

'accept-language': 'zh-CN,zh;q=0.9',

'referer': 'https://you.ctrip.com/travels/ningxia100063/t3-p3.html',

'sec-ch-ua': '"Not:A-Brand";v="99", "Chromium";v="112"',

'sec-ch-ua-mobile': '?0',

'sec-ch-ua-platform': '"Windows"',

'sec-fetch-dest': 'document',

'sec-fetch-mode': 'navigate',

'sec-fetch-site': 'same-origin',

'sec-fetch-user': '?1',

'upgrade-insecure-requests': '1',

'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/112.0.0.0 Safari/537.36 QuarkPC/2.0.7.222',

'x-uctiming-46938875': '1740111608630',

}

cookies = {

'GUID': '09031042116374938740',

'nfes\_isSupportWebP': '1',

'nfes\_isSupportWebP': '1',

'UBT\_VID': '1739436692970.debePwHkpldX',

'MKT\_CKID': '1739436696409.0u8jf.75s7',

'\_RSG': 'Bz4Z6rTUNYEFIxstuRSN69',

'\_RDG': '2887d7cc08b249256e01eb91092e3c7aa8',

'\_RGUID': 'e89779dd-468c-4f4b-9197-71be37dbb48f',

'\_bfaStatusPVSend': '1',

'Hm\_lvt\_e4211314613fcf074540918eb10eeecb': '1739436693,1740110042',

'HMACCOUNT': 'D1D7F8ADA1F46996',

'ASP.NET\_SessionSvc': 'MTAuNjEuMjIuMjUxfDkwOTB8amlucWlhb3xkZWZhdWx0fDE3MDUwMzE4MzY0NjQ',

'\_RF1': '222.89.92.90',

'\_ubtstatus': '%7B%22vid%22%3A%221739436692970.debePwHkpldX%22%2C%22sid%22%3A2%2C%22pvid%22%3A9%2C%22pid%22%3A290570%7D',

'\_bfi': 'p1%3D290570%26p2%3D290570%26v1%3D9%26v2%3D7',

'\_bfaStatus': 'success',

'\_pd': '%7B%22\_o%22%3A33%2C%22s%22%3A224%2C%22\_s%22%3A4%7D',

'Hm\_lpvt\_e4211314613fcf074540918eb10eeecb': '1740111561',

'Hm\_lvt\_a8d6737197d542432f4ff4abc6e06384': '1740111574',

'Hm\_lpvt\_a8d6737197d542432f4ff4abc6e06384': '1740111574',

'HMACCOUNT': 'D1D7F8ADA1F46996',

'\_bfa': '1.1739436692970.debePwHkpldX.1.1740111571781.1740111573947.2.12.102001',

'\_ga': 'GA1.1.381522301.1740111575',

'\_ga\_9BZF483VNQ': 'GS1.1.1740111575.1.0.1740111575.0.0.0',

'\_ga\_5DVRDQD429': 'GS1.1.1740111575.1.0.1740111576.0.0.0',

'\_ga\_B77BES1Z8Z': 'GS1.1.1740111575.1.0.1740111576.59.0.0',

'\_jzqco': '%7C%7C%7C%7C1740110051002%7C1.1022539780.1739436696415.1740111570726.1740111578066.1740111570726.1740111578066.undefined.0.0.12.12',

'MKT\_Pagesource': 'PC',

}

# æ‰“å¼€CSVæ–‡ä»¶å¹¶åˆå§‹åŒ–writer

with open('æºç¨‹.csv', 'w', newline='', encoding='utf-8-sig') as csvfile:

fieldnames = ['æ ‡é¢˜', 'å‘å¸ƒæ—¶é—´', 'æ­£æ–‡']

writer = csv.DictWriter(csvfile, fieldnames=fieldnames)

writer.writeheader()

# éåŽ†æ¯ä¸€é¡µ

for i in range(1, 11):

url = f'https://you.ctrip.com/travels/ningxia100063/t3-p{i}.html'

response = requests.get(url, cookies=cookies, headers=headers)

tree = etree.HTML(response.text)

# èŽ·å–æ¯ä¸ªæ–‡ç« çš„é“¾æŽ¥

articles = tree.xpath('//\*[@id="content"]/div[4]/div/div[2]/div/div[2]/a')

for article in articles:

title = article.xpath('.//div/dl/dt/text()')[0]

times = article.xpath('.//div/dl/dd[1]/text()')[0].replace('\xa0', '').strip()

ahref = article.attrib['href']

ahref = 'https://you.ctrip.com' + ahref

# è¯·æ±‚æ–‡ç« è¯¦æƒ…é¡µ

res = requests.get(ahref, cookies=cookies, headers=headers)

tree = etree.HTML(res.text)

# æå–æ–‡ç« æ­£æ–‡

ptext = tree.xpath('//div[@class="ctd\_content"]//p/text()')

text = ''.join(ptext).replace('\xa0', '').replace('\n', '').replace('\r', '').strip()

# å†™å…¥CSV

dit = {'æ ‡é¢˜': title, 'å‘å¸ƒæ—¶é—´': times, 'æ­£æ–‡': text}

print(dit)

writer.writerow(dit)