

```
# -*- coding: utf-8 -*-
# @Author: Cheung Y.H.
# @Date : 2021/11/11
# version: Python 3.9.2
'''
```

作者敬告：本代码仅用于学术交流，作者不负版权责任。请勿从事任何违法活动！

请访问出版社下载阅读正版文章！

```
Author Warning: This code is only used for academic communication.
                 The author has no liability for copyright.
                 DO NOT ENGAGE IN ANY ILLEGAL ACTIVITIES.
                 Please download and read the genuine articles from the publisher.
'''
```

```
import requests
import time
from bs4 import BeautifulSoup
import os
```

```
path = "D:\\userdata\\download\\Pb\\" #下载后的 PDF 存放的文件夹
```

```
if os.path.exists(path) == False:
```

```
    os.mkdir(path) #20210607 更新，创建保存下载文章的文件夹
```

```
f = open("D:\\userdata\\download\\Pb.txt", "r", encoding="utf-8-sig") #存放 DOI 码
```

的.txt 文件中，每行存放一个文献的 DOI 码，完毕须换行（最后一个也须换行！）

```
head = {\
```

```
        'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64)
        AppleWebKit/537.36 (KHTML, like Gecko) Chrome/79.0.3945.117 Safari/537.36\'
```

```
    } #20210607 更新，防止 HTTP403 错误
```

```
for line in f.readlines():
```

```
    line = line[:-1] #去换行符
```

```
    #url = "https://www.sci-hub.pl/" + line + "#" #sci hub 的 DOI 码检索 url 这个
地址已经失效了
```

```
    url = "https://www.sci-hub.ren/" + line + "#" #20210515 更新：现在换成这个 sci
hub 检索地址
```

```

time.sleep(10) #2022/9/14
try:
    download_url = "" #20211111 更新

    r = requests.get(url, headers = head)
    r.raise_for_status()
    r.encoding = r.apparent_encoding
    soup = BeautifulSoup(r.text, "html.parser")
    #download_url = "https:" + str(soup.div.ul.find(href="#")).split("href=")[-
1].split(".pdf")[0] + ".pdf" #找出存放该文献的 pdf 资源地址 (该检索已失效)

    if soup.iframe == None: #20211111 更新

        download_url = "https:" + soup.embed.attrs["src"] #20211111 更新
    else:
        download_url = soup.iframe.attrs["src"] #20210515 更新

    print(line + " is downloading...\n --The download url is: " + download_url)
    download_r = requests.get(download_url, headers = head)
    download_r.raise_for_status()
    with open(path + line.replace("/", "_") + ".pdf", "wb+") as temp:
        temp.write(download_r.content)
except:
    with open("Pberror.txt", "a+") as error:
        error.write(line + "\n")
        if "https://" in download_url:
            error.write(" --The download url is: " + download_url + "\n\n")
else:
    download_url = "" #20210801 更新

    print(line + " download successfully.\n")
f.close()

```