

Task 1 ReadMe

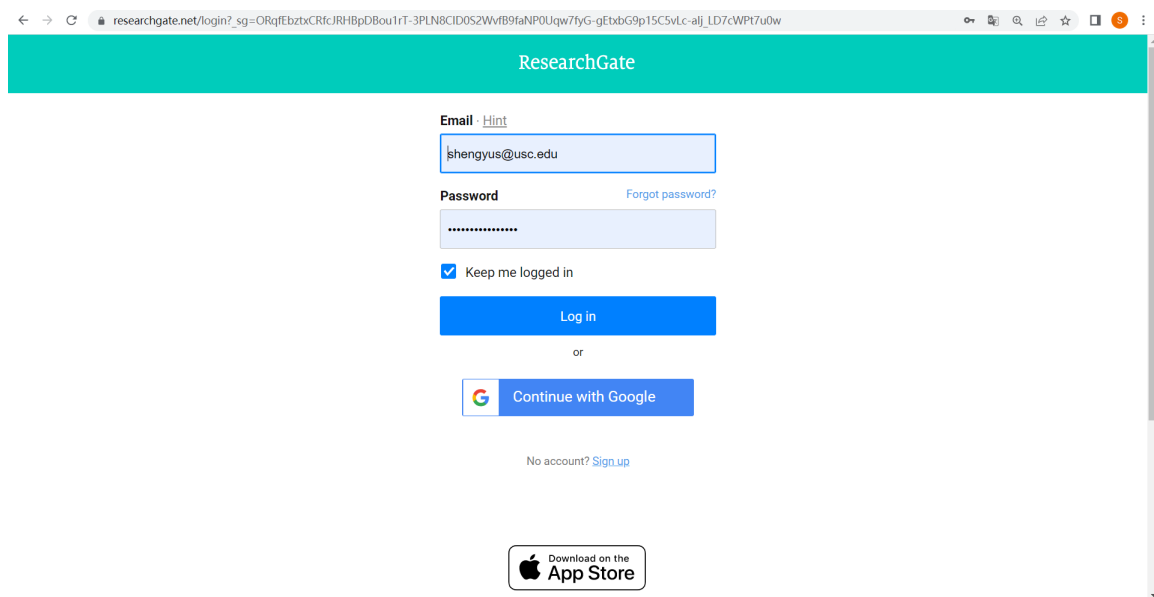
This part is used to crawl the pdf files of papers in bik dataset.

After homework1, we found that there is no single website for storing all required paper PDF files. Therefore, in this crawling task, we selected the following two websites:

Research Gate: <https://www.researchgate.net/>
USC library: <https://libraries.usc.edu/>

Since the member who do the web crawling task is not in the United States, he need VPN to access the USC library. However, VPN will conflict with the requests library. Therefore, we chose to use selenium to simulate browser access and download for crawling.

The code below shows our first step: simulate the browser to query and crawl the Research Gate. To crawl this website, we need to login first, the user_email and user_pwd variables are the exact account information.



After login at the Research Gate, we can search the papers by DOI number. By simulating the browser, we can click the Download button and finish crawling.

researchgate.net/publication/237057984_HIV-1_Tat_Protein_Increases_Microglial_Outward_K⁺_Current_and_Resultant_Neurotoxic_Activity

Article Full-text available Funding Information

HIV-1 Tat Protein Increases Microglial Outward K⁺ Current and Resultant Neurotoxic Activity

May 2013 · PLoS ONE 8(5):e64904 · [Follow journal](#)
 DOI: 10.1371/journal.pone.0064904
 Source: PubMed
 License: CC BY 4.0

Jianuo Liu · Peng Xu · Cory Collins · [Show all 7 authors](#) · Huangui Xiong

Research Interest 13.6
 Citations 19
 Recommendations 0
 Reads 19 new 442
[See details](#)

Overview Stats Comments Citations (19) References (72) ... [Download](#) Share More

Abstract and figures

Microglia plays a crucial role in the pathogenesis of HIV-1-associated neurocognitive disorders. Increasing evidence indicates the voltage-gated potassium (Kv) channels are involved in the regulation of microglia function, prompting us to hypothesize Kv channels may also be involved in microglia-mediated neurotoxic activity in HIV-1-infected brain. To test this hypothesis, we investigated the involvement of Kv channels in the response of microglia to HIV-1 Tat protein. Treatment of rat microglia with HIV-1 Tat protein (200 ng/ml) resulted in pro-inflammatory microglial activation, as indicated by increases in TNF- α , IL-1 β , reactive oxygen species, and nitric oxide, which were

Royal Society Publishing
Proceedings B 2022 Ecology collection
 Submissions welcome before 31 July 2022

Call for papers
 Contribute to the new Proceedings B Ecology collection

```
river=webdriver.Chrome()
driver.get("https://www.researchgate.net/login")
driver.find_element(By.ID, "input-login").send_keys(user_email)
driver.find_element(By.ID, "input-password").send_keys(user_pwd)
driver.find_element(By.CLASS_NAME, "nova-legacy-c-button__label").find_element(By.XPATH, "../..").click()
for i in range(165, len(bik)):
    curr_page = 'https://www.researchgate.net/search/publication?q='+bik['DOI']
    [i]
    driver.get(curr_page)
    bs_test=bs(driver.page_source, 'html.parser')
    if bs_test.find('a', {'class': 'nova-legacy-c-button nova-legacy-c-button--align-center nova-legacy-c-button--radius-m nova-legacy-c-button--size-m nova-legacy-c-button--color-blue nova-legacy-c-button--theme-solid nova-legacy-c-button--width-auto'})!=None:
        driver.find_element(By.LINK_TEXT, "Download").click()
```

However, when the search result is Request full-text instead of Download, we need to find another way to crawl.

researchgate.net/publication/306007080_A_Critical_Role_for_Notch_Signaling_in_the_Formation_of_Cholangiocellular_Carcinomas

Article

A Critical Role for Notch Signaling in the Formation of Cholangiocellular Carcinomas

August 2016 · Cancer Cell 30(2):353-356 · [Follow journal](#)
 DOI: 10.1016/j.ccell.2016.07.005

Steffen Zender · Irina Nickleleit · Torsten Wuestefeld · [Show all 15 authors](#) · Nisar P. Malek

Research Interest 14.6
 Citations 27
 Recommendations 0
 Reads 1 new 87
[See details](#)

Overview Stats Comments Citations (27) References (1) ... [Request full-text](#) Share More

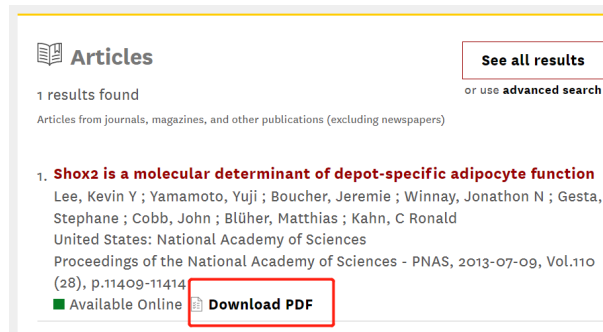
Abstract

(Cancer Cell 23, 784–795; June 10, 2013) In the original article, which shows that the expression of Notch-ICD leads to the formation of intrahepatic cholangiocarcinomas, there is a duplication of actin blots in Figures 2G, 6F, and S2F. Figure 2G shows the

InvivoGen

Recommend Recommend this work Follow Get updates Share Share in a message

There are still 100+ papers could not find in Research Gate, so we access USC library to finish the crawling task of the remain papers. The code below shows the process of downloading pdf files from USC library by simulating Chrome browser as well.



```
else:
    curr_page = 'https://www.researchgate.net/search/publication?
q='+bik['Title'][i]
    driver.get(curr_page)
    bs_title=bs(driver.page_source,'html.parser')
    if bs_title.find('a',{'class':'nova-legacy-c-button nova-legacy-c-
button--align-center nova-legacy-c-button--radius-m nova-legacy-c-button--size-s
nova-legacy-c-button--color-blue nova-legacy-c-button--theme-ghost nova-legacy-
c-button--width-auto'})!=None:
        url=bs_title.find('a',{'class':'nova-legacy-c-button nova-legacy-c-
button--align-center nova-legacy-c-button--radius-m nova-legacy-c-button--size-s
nova-legacy-c-button--color-blue nova-legacy-c-button--theme-ghost nova-legacy-
c-button--width-auto'})['href']
        pdf_url='https://www.researchgate.net/'+url
        driver.find_element(By.LINK_TEXT, "Download").click()
    print('finish '+str(i))
```

After these two steps, we can collect 200+ papers mentioned in the bik dataset.