

Task Report

Task: To scrape links from Whatsapp chats/groups

What is selenium?

Selenium with Python is used to carry out automated test cases for browsers or web applications. You can easily use it to simulate tests such as tapping on a button, entering content to the structures, skimming the entire site, etc

Requirements: python ide, selenium package

Steps:

- 1) Go to command prompt, run it as administrator. Type "pip install selenium". Your package will be installed
- 2) Go to python IDE. I am using jupyter notebook and type the code given below and run it

Code:

```
# Python program to demonstrate
# selenium

# import webdriver
from selenium import webdriver
import time
from selenium.webdriver.common.keys import Keys
import requests
from selenium.webdriver.common.by import By
import os

chat_name=input("please enter the name of the chat you want to find the links from")
# create webdriver object
driver = webdriver.Chrome()

# get google.co.in
driver.get("https://web.whatsapp.com/")
time.sleep(35)
search_box = driver.find_element(By.XPATH,'//div[@contenteditable="true"]')
search_box.send_keys(chat_name)
time.sleep(2)
search_box.send_keys(Keys.ENTER)
time.sleep(60)

for i in range(30):
    driver.execute_script("document.body.scrollTop += 500;")
    time.sleep(20)

message_bubbles = driver.find_elements(By.XPATH, '//div[contains(@class, "message-in") or contains(@class, "message-out")]')
import re
# Extract and filter HTTPS links using regular expressions
https_links = []
for bubble in message_bubbles:
    message = bubble.text
    links = re.findall(r'https://[^\s]+', message)
    https_links.extend(links)

# Print the extracted links
for link in https_links:
    print(link)

driver.quit()
```

Output:

please enter the name of the chat you want to find the links fromBCA Alerts!!

<https://docs.google.com/forms/d/e/1FAIpq>

<https://docs.google.com/forms/d/e/1FAIpq>

<https://flipped.ai/org/flipped/jobs/FSD>

<https://flipped.ai/org/dashboard/jobs/F>

<https://docs.google.com/forms/d/e/1FAIpq>

<https://docs.google.com/forms/d/e/1FAIpq>

<https://docs.google.com/forms/d/e/1FAIpq>

<https://docs.google.com/forms/d/e/1FAIpq>

Note: This code can work on both group chats and personal chats