

LLM, RAG – Web Scrapping

1. Search websites for offensive cyber data.

Websites such as :

- Exploit-db
- Huggingface
- OWASP

2. Run code into Jupyter Notebook

```
import os
import time
from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC

download_dir = "/home/ubuntu/Downloads/cyber_data"
base_url = "https://www.exploit-db.com/download/"

# Function to download a file given the download URL
def download_file(url, driver):
    driver.get(url) # Load the download URL in the browser
    time.sleep(2) # Give some time for the download to start

# Main function to iterate through all pages
def main():
    driver = webdriver.Chrome()

    try:
        for num in range(52060, 0, -1): # Range from 1 to 52060
            url = f"{base_url}{num}"
            print(f"Downloading file from: {url}")

            download_file(url, driver)
            time.sleep(2) # Add a short delay between downloads

    except Exception as e:
        print(f"Error: {e}")

    finally:
        driver.quit() # Make sure to quit the WebDriver session at the end

if __name__ == "__main__":
    if not os.path.exists(download_dir):
        os.makedirs(download_dir)
    main()
```

3. Receive results from configuration commands:

[illegible]