LLM, RAG - Web Scraping

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()
       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}")
       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)
```

3. Receive results from configuration commands:

```
Downloading file from: https://www.exploit-db.com/download/52060
Downloading file from: https://www.exploit-db.com/download/52059
Downloading file from: https://www.exploit-db.com/download/52058
Downloading file from: https://www.exploit-db.com/download/52057
Downloading file from: https://www.exploit-db.com/download/52056
Downloading file from: https://www.exploit-db.com/download/52055
Downloading file from: https://www.exploit-db.com/download/52054
Downloading file from: https://www.exploit-db.com/download/52053
Downloading file from: https://www.exploit-db.com/download/52052
Downloading file from: https://www.exploit-db.com/download/52051
Downloading file from: https://www.exploit-db.com/download/52050
Downloading file from: https://www.exploit-db.com/download/52049
Downloading file from: https://www.exploit-db.com/download/52048
Downloading file from: https://www.exploit-db.com/download/52047
Downloading file from: https://www.exploit-db.com/download/52046
Downloading file from: https://www.exploit-db.com/download/52045
Downloading file from: https://www.exploit-db.com/download/52044
Downloading file from: https://www.exploit-db.com/download/52043
Downloading file from: https://www.exploit-db.com/download/52042
Downloading file from: https://www.exploit-db.com/download/52041
Downloading file from: https://www.exploit-db.com/download/52040
Downloading file from: https://www.exploit-db.com/download/52039
Downloading file from: https://www.exploit-db.com/download/52038
Downloading file from: https://www.exploit-db.com/download/52037
Downloading file from: https://www.exploit-db.com/download/52036
Downloading file from: https://www.exploit-db.com/download/52035
Downloading file from: https://www.exploit-db.com/download/52034
Downloading file from: https://www.exploit-db.com/download/52033
Downloading file from: https://www.exploit-db.com/download/52032
Downloading file from: https://www.exploit-db.com/download/52031
Downloading file from: https://www.exploit-db.com/download/52030
Downloading file from: https://www.exploit-db.com/download/52029
Downloading file from: https://www.exploit-db.com/download/52028
Downloading file from: https://www.exploit-db.com/download/52027
Downloading file from: https://www.exploit-db.com/download/52026
Downloading file from: https://www.exploit-db.com/download/52025
Downloading file from: https://www.exploit-db.com/download/52024
Downloading file from: https://www.exploit-db.com/download/52023
Downloading file from: https://www.exploit-db.com/download/52022
```