

Forum:[Hack Forums](#)

Thread: [Web Scraper \(AI\) script Builder?](#)

devound - August 19, 2024, 7:27 pm

Is there any tool or ChatGPT prompt to easily generate web scraping scripts? I want a script to iterate crypto projects and grab each discord and telegram join link

Ecca - August 19, 2024, 10:08 pm

You don't need a jailbreak, [just ask ChatGPT like this](#).

With a little editing this works, the github query for crypto projects needs to be adjusted though:

Code import requests

from bs4 import BeautifulSoup

import re

import base64

GitHub API details

GITHUB_API_URL = "https://api.github.com"

GITHUB_TOKEN = "{github_token}" # Replace with your GitHub personal access token

HEADERS = {"Authorization": f"token {GITHUB_TOKEN}"}

Search for crypto projects on GitHub

def search_github_projects(query, max_results=10):

 search_url = f"{GITHUB_API_URL}/search/repositories?q={query}&sort=stars&order=desc"

 response = requests.get(search_url, headers=HEADERS)

 if response.status_code == 200:

 data = response.json()

 return data['items'][:max_results]

 else:

 print(f"Failed to fetch projects: {response.status_code}")

 return []

Extract URLs from README

def extract_urls_from_readme(owner, repo):

 readme_url = f"{GITHUB_API_URL}/repos/{owner}/{repo}/readme"

 response = requests.get(readme_url, headers=HEADERS)

 if response.status_code == 200:

 readme_data = response.json()

 readme_content = base64.b64decode(readme_data['content']).decode('utf-8')

 urls = re.findall(r'(https?://\S+)', readme_content)

 return [url for url in urls if "#" in url.lower()

 or "discordapp.com" in url.lower()

 or "t.me" in url.lower()

 or "telegram.org" in url.lower()]

 else:

 print(f"Failed to fetch README for {owner}/{repo}: {response.status_code}")

 return []

Main function to scrape Discord and Telegram URLs

```
def main():
    f = open("web_scaper_output.txt", "a")
    query = "crypto"
    projects = search_github_projects(query)
    for project in projects:
        owner = project['owner']['login']
        repo = project['name']
        print(f"Scraping {owner}/{repo}...")
        urls = extract_urls_from_readme(owner, repo)
        for url in urls:
            f.write(url + "\n")
    f.close()

if __name__ == "__main__":
    main()
```

Klonopin - August 21, 2024, 2:54 pm

AI won't be able to generate you the xpaths