PYTHON PROJECT

DI CHƠI ĐÂU ??

GROUP 5

Vũ Minh An Chu Cao Nguyên Phạm Tiến Dũng Trịnh Mai Anh Nguyễn Khánh Huyền Nhan Yến Trang

Crawl data from google chrome and facebook

OUTLINE

Main modules

3

Deploy online web

Craw Data

Crawl Data

Selenium

- Automation test tools (kiểm thử phần mềm hoặc website tự động)
- Mô phỏng hầu hết những thao tác người dùng với trình duyệt (nhấp – click, cuộn – scroll, đóng tab, click check box, ...)

webdriver.Chrome(service = Service ("./chromedriver.exe"))

- Đây là nơi yêu cầu Selenium sử dụng để tạo ra một phiên bản Google Chrome mới
- Giá trị './chromedriver.exe' mà chúng ta đang chuyển tới phương thức .Chrome () là vị trí của tệp chromedriver mà chúng ta đã tải xuống.

browser = webdriver.Chrome(service =
 Service("./chromedriver.exe"))

browser.get ('https://www.google.com')

- Trình điều khiển Selenium điều hướng đến URL
- 'https://www.google.com' : URL

browser.find_element(By.CLASS_NAME, '..')

=> first element match the class name

browser.find_elements(By.XPATH, '..')

=> List elements match the xpath

CRAWL DATA BẰNG MỘT SỐ PHƯƠNG THỰC SẪN CÓ CỦA SELENIUM

browser.find_element(By.LINK_TEXT, '..')

=> first element match the link textt

element.click()

=>để click vào element được chỉ định

Crawl Data

http.clients

 thư viện giúp truy cập web và crawl dữ liệu bằng cách truy vấn api của web

http.client.HTTPSConnection(link)

truy cập đến đường link cho trước

http.client.request()

• truy vấn api của web

http.client.HTTPSConnection.getresponse()

• lấy phản hồi (data) từ api web

.read()

load data sau khi đã getrespone

.decode("utf-8")

đưa data đã crawl về sang dạng string

MAIN MODULES

STREAMLIT

Thư viện cho phép trình bày data và machine learning thành một trang web có sự tương tác, giao diện.

HÀM SỬ DỤNG

- **streamlit.dataframe()**: Đưa bảng dữ liệu lên trên streamlit web
- **streamlit.header()**: viết tiêu đề (header)
- **streamlit.write()**: viết thông tin lên streamlit web
- streamlit.sidebar.selectbox(box_name[list _of_option]): Tạo 1 box gồm các lựa chọn cho người dùng

VÍ DỤ

streamlit.dataframe(df_event, None, 500)

streamlit.header(f'Ha Noi {end point}')
streamlit.write("###")

streamlit.sidebar.selectbox('Choice', ['Game', 'Event', 'Film'])

STREAMLIT

Thư viện cho phép trình bày data và machine learning thành một trang web có sự tương tác, giao diện.

HÀM SỬ DỤNG

- **streamlit.container()**: tạo ô chứa
- **streamlit.columns(number)**: tạo ra số hàng
- **streamlit.date_input()**: Tạo một ô cho phép người dùng chọn thời gian cụ thể 1 ngày
- **streamlit.sidebar.header()**: Tiêu đề trên thanh sidebar (bên trái)
- streamlit.image(): hiển thị ảnh lên streamlit

VÍ DỤ

with streamlit.container(): col1, col2 = streamlit.columns(2)

streamlit.date_input("Which day that you wanna join an event?", datetime.date(2022,11,08))

streamlit.sidebar.header('Welcome to Đi chơi đâu team web')

streamlit.image(game_key[i].picture[0])

CÁC THƯ VIỆN PHỤ CỦA STREAMLIT

Các thư viện bổ trợ cho việc hiển thị dữ liệu trên streamlit

streamlit_folium

Thư viện giúp trình bày bản đồ lên streamlit web

from streamlit_folium import st_folium

st_folium(map, with = 700, height = 450)
Hiện bản đồ lên với thông số chiều dài là 700 và chiều cao 450

streamlit_lottie

Thư viện giúp đẩy ảnh động lên streamlit web

from streamlit_lottie import st_lottie

st_folium(spaceship, height = 200, key = 'spaceship') Câu lệnh cho load ảnh động lên streamlit web

SOME ANOTHER MODULES

OPENPYXL

Ví dụ

Thư viện cho phép truy cập các thông tin qua ứng dụng Excel:

openpyxl.load_book(): đây là một hàm để load file excel thông qua đường dẫn wb = load_workbook(r'data1/DataFrame.xlsx')ws = wb.active

=>đây là một hàm để load file excel thông qua đường dẫn

Ví dụ

- datetime.datetime.now()
 - => lấy thời gian bây giờ
- datetime.datetime.now() + datetime.timedelta(days=delta)
 - =>ngày tương lai = ngày hôm nay + chênh lệch số ngày (delta)
- datetime.datetime.strptime(strn[0], '%a, %d %b %Y')
 - =>convert string với dạng cho trước (ở đây là thứ, ngày tháng năm) về type datetime
- coming_date.strftime("%d/%m/%Y")
 - => đưa từ type datetime sang type string

Module cung cấp các hàm xử lí thời gian

Crawl: convert dạng thời gian về chung 1 kiểu hiển thị

Module: làm filter cho người dùng chọn thời gian

DATETIME

FOLIUM

Thư viện cho phép người dùng sử dụng bản đồ google và có thể đánh dấu hoặc bôi các vị trí trên bản đồ thông qua kinh tuyến vĩ tuyến

Ví dụ

- map = folium.Map(location= [a,b] ,zoom_start =11)
- => khởi tạo google map tại vị trí kinh tuyến a, vĩ tuyến b với khoảng cách zoom bắt đầu là 11
- folium.Marker(location =[game_key[id].lattitude, game_key[id] longtitude])
- =>Hàm này giúp đánh dấu vị trí tại kinh tuyến vĩ tuyến và và có thể cho thêm thông tin đi kèm khi bấm vào vị trí đó

Ví dụ

- data_df = pd.DataFrame(data = data_format)=> Tao data frame
- df_event = pd.read_csv('data1/events.csv')=> Đọc file csv

Pandas được thiết kế để làm việc dễ dàng và trực quan với dữ liệu có cấu trúc

Crawl: Tạo data frame

Module: đọc file csv

PANDAS

FORMSUBMIT

Trang web giúp có thể đưa 1 dạng contact form vào trong web thông qua html

- Đoạn code bao gồm 3 ô điền name, text, email và bên trong ô có chỉ dẫn cho người dùng (Your name, your email, your message here)
- Một nút Submit hiện thị là ô Send trên streamlit web

DEPLOY ONLINE WEB

vuxminhan Rename main.py to app.py X

Ax 1 contributor

```
273 lines (240 sloc) 9.47 KB
      import streamlit as st
      from openpyxl import load_workbook
      import pandas as pd
       import folium
      from streamlit_folium import st_folium
      import random as rd
       import time
       import datetime
       import requests
      from streamlit_lottie import st_lottie
 12
 13
      # get animation
      def load lottieurl(url):
          r = requests.get(url)
          if r.status_code |- 200:
```

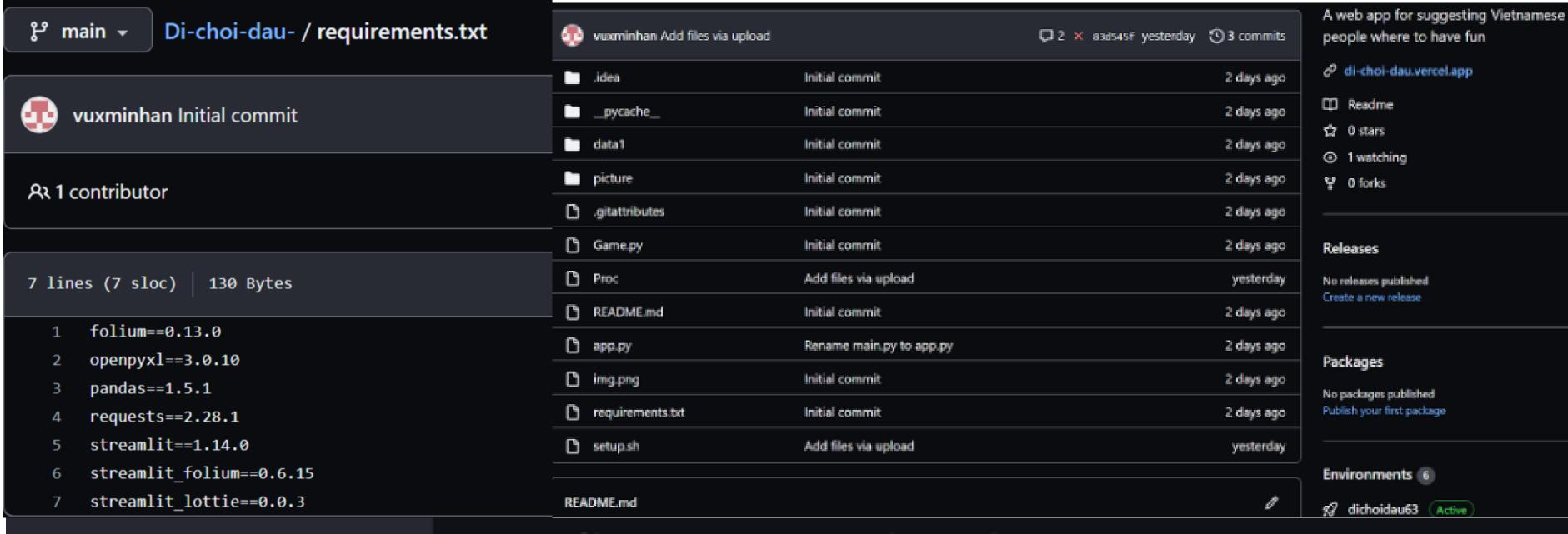
STREAMLIT

- Tạo App's Script: app.py
- => Giúp streamlit cloud biết cần hiển thị những gì trên server
 - Tạo file proc để setup và requirement.txt
 - => để server tải đủ các library cần thiết

GITHUB

- Tạo Public GitHub Repository
- => để streamlit cloud có đủ tất cả files cần để chạy web
 - Két női repo github với Streamlit Cloud
 - Deploy trên Streamlit cloud

Method



Filter

Another Place

Welcome to Di choi dau team web

Choice

Game

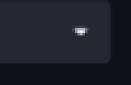
Some sidebar text

Number of rating is higher Uhami

Rating is higher than

10

2.0



Choose range time opening



