**Crwal data Elise**

1. import các thư viện và module:

from selenium import webdriver: Import module webdriver từ thư viện Selenium để tương tác với trình duyệt.

from selenium.webdriver.common.by import By: Import class By để sử dụng trong việc xác định vị trí các phần tử.

from bs4 import BeautifulSoup: Import class BeautifulSoup từ thư viện BeautifulSoup để parse HTML.

import pandas as pd: Import thư viện pandas với alias là pd để làm việc với dữ liệu dưới dạng DataFrame.

import os: Import module os để thao tác với hệ điều hành.

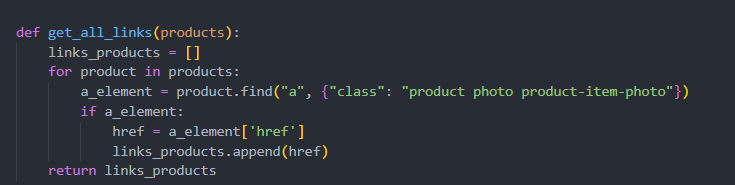
from selenium.webdriver.support.ui import WebDriverWait: Import class WebDriverWait để thực hiện các chờ đợi thông minh trong Selenium.

from selenium.webdriver.support import expected\_conditions as EC: Import module expected\_conditions với alias là EC để sử dụng các điều kiện chờ đợi trong Selenium.

from selenium.webdriver.common.action\_chains import ActionChains: Import class ActionChains để thực hiện các hành động chuỗi trong Selenium.

import time: Import module time để sử dụng các hàm liên quan đến thời gian.

1. Hàm get\_all\_links:



Hàm này nhận vào danh sách products và trả về danh sách các đường link (href) của các sản phẩm.

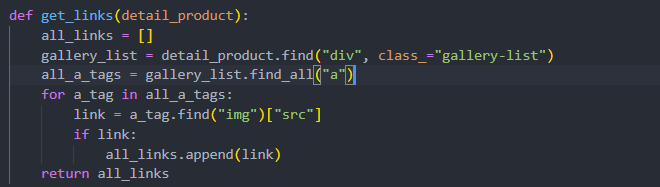
Sử dụng vòng lặp để duyệt qua từng sản phẩm.

Sử dụng phương thức find của BeautifulSoup để tìm phần tử a có class là "product photo product-item-photo".

Nếu phần tử a được tìm thấy, lấy giá trị của thuộc tính href và thêm vào danh sách links\_products.

Trả về danh sách các đường link.

1. Hàm get\_links:



Hàm này nhận vào chi tiết sản phẩm detail\_product và trả về danh sách các đường link của ảnh sản phẩm.

Tìm phần tử div có class là "gallery-list" trong detail\_product.

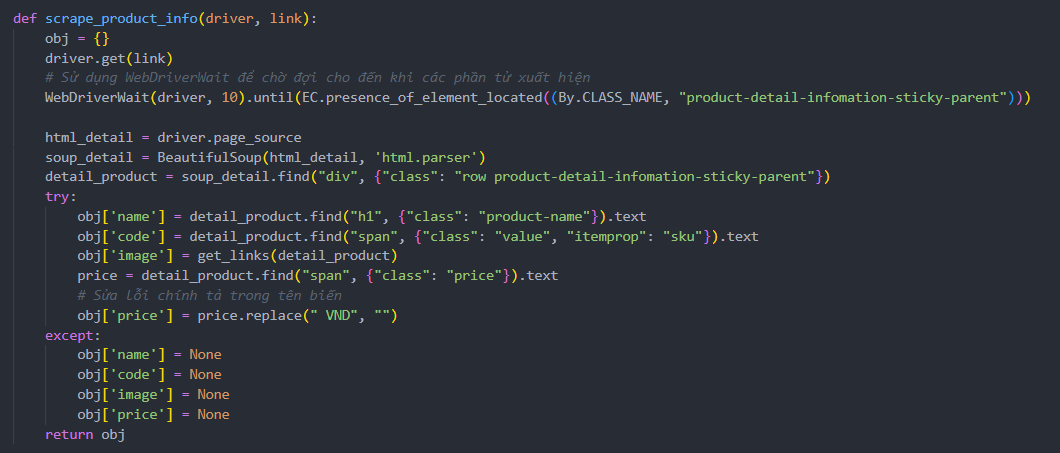
Tìm tất cả các phần tử a trong gallery\_list.

Duyệt qua từng phần tử a và lấy giá trị của thuộc tính src từ phần tử img nếu tồn tại.

Thêm các đường link vào danh sách all\_links.

Trả về danh sách các đường link.

1. Hàm scrape\_product\_info:



Hàm này nhận vào trình duyệt driver và đường link link của sản phẩm, sau đó trả về một đối tượng chứa thông tin sản phẩm.

Mở trình duyệt với đường link của sản phẩm.

Sử dụng WebDriverWait để chờ đợi cho đến khi phần tử có class "product-detail-infomation-sticky-parent" xuất hiện.

Lấy mã nguồn HTML của trang và sử dụng BeautifulSoup để parse.

Tìm phần tử div có class là "row product-detail-infomation-sticky-parent" để lấy thông tin chi tiết sản phẩm.

Sử dụng try và except để bắt lỗi trong quá trình lấy thông tin sản phẩm.

Nếu không có lỗi, lấy thông tin như tên, mã, ảnh, và giá của sản phẩm.

Nếu có lỗi, gán các giá trị là None.

Trả về đối tượng chứa thông tin sản phẩm.

1. Hàm main:



Hàm chính của chương trình.

Khởi tạo một đối tượng trình duyệt (driver) và thiết lập kích thước cửa sổ trình duyệt.

Sử dụng vòng lặp để lặp qua các trang của website.

In ra thông báo về lần lặp hiện tại.

Mở trang web và chờ đợi cho đến khi phần tử có class "products" xuất hiện.

Lấy mã nguồn HTML của trang và sử dụng BeautifulSoup để parse.

Tìm tất cả các sản phẩm và gọi hàm get\_all\_links để lấy danh sách đường link của các sản phẩm.

Duyệt qua từng đường link, gọi hàm scrape\_product\_info để lấy thông tin chi tiết và in ra thông tin của sản phẩm.

Thêm thông tin của sản phẩm vào danh sách data.

Tạo DataFrame từ danh sách data.

Ghi dữ liệu vào file CSV (test.csv).

Kết thúc chương trình và đóng trình duyệt.