

Nama : Imas Nabellia Venda

NIM : 1123102116

Tugas 3 Selenium

A. Kode Python

```
from selenium import webdriver
from bs4 import BeautifulSoup

URL = 'https://www.bukalapak.com/products?'
params = {
    'D':'batik%20banyuwangi'
}
headers = {'User-Agent':'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/131.0.0.0 Safari/537.36'}
driver = webdriver.Chrome()
fullURL = f"{URL}&search%5Bkeywords%5D={params['D']}"
driver.get(fullURL)
html = driver.page_source

soup = BeautifulSoup(html, 'html.parser')
data = soup.find_all('div',{'class':'bl-flex-item mb-8'})

for i in range(len(data)):
    nama = data[i].find('p', {'class':'bl-text--secondary'})
    harga = data[i].find('p', {'class':'bl-product-card-new__price'})

    if nama and harga:
        print("Nama produk: " + nama.text.strip())
        print("Harga: Rp." + harga.text.strip())

    link_produk = data[i].find('a', {'href':True})
    if link_produk:
        link = link_produk['href']
        driver.get(link)
        html_detail = driver.page_source
        soup_detail = BeautifulSoup(html_detail, 'html.parser')

        penjual = soup_detail.find('h3', {'class':'c-seller__name'})
        print("Penjual: " + penjual.text, "\n")

    driver.back()

driver.quit()
```

B. Penjelasan Kode

- Import library Selenium Webdriver dan BeautifulSoup.
- Ambil URL dari website yang ingin diambil datanya.
- Parameter digunakan untuk kata kunci pencarian. Di sini saya mencari produk dengan keyword 'batik banyuwangi'. Nilai parameter dapat diketahui pada link setelah kita melakukan pencarian.
- Header digunakan untuk menentukan User-Agent. User-Agent dapat diketahui dengan menklik inspeksi pada halaman web. User-Agent terletak pada bagian network.
- Program akan membuka Chrome menggunakan Selenium.
- Untuk membentuk URL penuh didapat dari penggabungan URL awal dan kata kunci pencarian. URL yang sudah lengkap akan dibuka.
- Program akan mengambil konten HTML dan diparsing menggunakan BeautifulSoup. Lalu akan melakukan pencarian semua elemen <div> yang memiliki class bl-flex-item mb-8. Class tersebut adalah container untuk setiap produk yang ada di halaman web.
- Iterasi setiap elemen, looping akan berjalan sebanyak jumlah elemen yang ada dalam data. Nama produk didapat melalui pencarian <p> dengan class bl-text—secondary. Harga produk dicari berdasarkan <p> dengan class bl-product-card-new__price. Saat menentukan class harus sesuai dengan class yang ada di struktur kode website.
- Jika kedua nilai dari nama dan harga tidak kosong akan menampilkan nilai tersebut dilayar.
- Link-produk akan mengambil detail link dan penjual. Pencarian akan dilakukan pada elemen <a> yang memiliki atribut href. Ini bertujuan untuk membuka halaman detail produk.
- Program akan mengambil HTML halaman detail produk dan akan diparsing menggunakan BeautifulSoup. Akan dilakukan pencarian pada elemen <h3> dengan class c-seller__name untuk mendapatkan nama penjual dan nama penjual akan ditampilkan.
- Setelah selesai mendapatkan nama penjual browser akan kembali pada halaman daftar produk.
- Browser akan ditutup untuk mengakhiri sesi Selenium.

C. Hasil

```
DevTools listening on ws://127.0.0.1:63159/devtools/browser/ed657045-d380-4c9f-bd5b-6a7b13706ae2
Nama produk: Baju KAOS Kerah Polo Shirt BATIK KHAS Daerah BANYUWANGI
Harga: Rp.120.000
Penjual: Ayung Sportindo

Nama produk: Batik Banyuwangi Ijo Lumut Kain Batik Pesisir Nusantara TokoShobi
Harga: Rp.125.000
Penjual: Toko Shobi

Nama produk: Baju batik Osing Banyuwangi
Harga: Rp.425.000
Penjual: Yonas Yonansyah

Nama produk: hijau || Iket Segitiga Banyuwangi Kuda Lumping Sleyer Ikat Kepala Udeng Batik Banyuwangi Jawa Timuranalf
Harga: Rp.21.000
Penjual: alfarizqy grosir surakarta

Nama produk: UNGU || Iket Segitiga Banyuwangi Kuda Lumping Sleyer Ikat Kepala Udeng Batik Banyuwangi Jawa Timuranalf
Harga: Rp.20.000
Penjual: alfarizqy grosir surakarta

Nama produk: BIRU || Iket Segitiga Banyuwangi Kuda Lumping Sleyer Ikat Kepala Udeng Batik Banyuwangi Jawa Timuranalf
Harga: Rp.20.000
Penjual: alfarizqy grosir surakarta
```

Contoh output

D. Analisa Hasil

Program berhasil mengambil nama produk, harga, dan penjual dari situs Bukalapak sesuai dengan kata kunci yaitu batik banyuwangi. Proses scraping berjalan dengan baik menggunakan Selenium untuk membuka halaman utama dan halaman detail produk. Navigasi bolak-balik antara halaman daftar produk dan halaman detail produk memerlukan waktu yang lama jika data yang diambil berjumlah banyak. Hasil yang didapatkan bergantung dari struktur HTML website tersebut.

E. Kesimpulan

Program telah berhasil mengekstrak data penting dari situs Bukalapak sesuai dengan kebutuhan. Namun, terdapat beberapa kelemahan yang memengaruhi efisiensi, seperti navigasi bolak-balik yang dapat dihindari dengan mengambil data langsung dari halaman utama jika memungkinkan.