

برای نوشتن اسکریپتی که به طور خودکار وبسایت‌ها را بازدید کرده و اطلاعات مورد نیاز را استخراج کند، می‌توانید از ابزارها و کتابخانه‌های مختلفی استفاده کنید. در اینجا دو روش رایج با استفاده از پایتون آورده شده است:

۱. استفاده از کتابخانه Beautiful Soup:

- **Beautiful Soup** یک کتابخانه محبوب پایتون برای تجزیه و تحلیل و استخراج اطلاعات از وبسایت‌ها است.
- با استفاده از این کتابخانه، می‌توانید به راحتی به محتوای HTML صفحات وب دسترسی پیدا کنید، عناصر خاصی را انتخاب کنید و داده‌های مورد نیاز خود را استخراج کنید.

مراحل:

۱. Beautiful Soup را نصب کنید:

Bash

```
pip install beautifulsoup4
```

۲. اسکریپت خود را بنویسید:

Python

```
import requests
from bs4 import BeautifulSoup

def scrape_website(url):
    # درخواست صفحه وب
    response = requests.get(url)

    # بررسی وضعیت پاسخ
    if response.status_code == 200:
        # محتوای HTML را به یک شی BeautifulSoup تبدیل کنید
        soup = BeautifulSoup(response.content, 'html.parser')

        # عنوان شرکت را پیدا کنید
        company_name = soup.find('h1', class_='company-name').text

        # اطلاعات تماس را پیدا کنید
        contact_info = soup.find('div', class_='contact-info')
        email = contact_info.find('a', href='mailto:').text
        phone_number = contact_info.find('span', class_='phone-number').text

        # محصولات را پیدا کنید
        products = []
```

```

for product_element in soup.find_all('div', class_='product'):
    product_name = product_element.find('h2').text
    product_price = product_element.find('span', class_='price').text
    products.append({
        'name': product_name,
        'price': product_price
    })

# اطلاعات استخراج شده را چاپ کنید
print(f"نام شرکت({company_name})")
print(f"ایمیل({email})")
print(f"شماره تلفن({phone_number})")
print(f"محصولات:")
for product in products:
    print(f" - {product['name']}: {product['price']}")

else:
    print(f"Error: Failed to load website. Status code: {response.status_code}")

if __name__ == '__main__':
url = "https://www-example-com.cdn.ampproject.org/c/s/www.example.com" #
URL وب سائیتی که می خواهید اسکرپ کنید را وارد کنید
scrape_website(url)

```

۲. استفاده از کتابخانه Selenium:

- Selenium ابزاری قدرتمندتر برای تعامل با مرورگرها و وبسایتها است.
- این کتابخانه به شما امکان می دهد تا اقداماتی مانند کلیک کردن روی دکمه ها، پر کردن فرم ها و پیمایش در صفحات را شبیه سازی کنید، که برای وبسایت هایی که با BeautifulSoup به راحتی قابل اسکرپیت نیستند مفید است.

مراحل:

۱. Selenium را نصب کنید:

```

bash
pip install selenium

```

۲. یک درایور مرورگر نصب کنید:

- برای استفاده از Selenium، باید درایور مرورگری را که می خواهید با آن تعامل داشته باشید نصب کنید.

○ برای مثال، برای استفاده از Chrome، باید درایور ChromeDriver را نصب کنید.

۳. اسکریپت خود را بنویسید:

```
Python
from selenium import webdriver
from selenium.webdriver.common.by import By

def scrape_website(url):
    # درایور Chrome را راه اندازی کنید
    driver = webdriver.Chrome()

    # به وب سایت بروید
    driver.get(url)

    # عنوان شرکت را پیدا کنید
    company_name_element = driver.find_element(By.CSS_SELECTOR,
                                                'h1.company-name')
    company_name = company_name_element.text

    # اطلاعات تماس را پیدا کنید
    contact_info_element = driver.find_element(By.CSS_SELECTOR, 'div.contact-
                                                info')
    email_element = contact_info_element.find_element(By.CSS_SELECTOR,
                                                       'a[href*="mailto:"]')
    email = email_element.text
    phone_number
```