

Web scraping

Веб-скрейпинг (или скрепинг, или скрапинг) — это технология получения веб-данных путем извлечения их со страниц веб-ресурсов.

Веб-скрейпинг объединяет в себе функции краулера (обход сайта и сбор данных) и парсера (анализ содержимого).



Где используется?

- Отслеживание цен на товары в интернет-магазинах и извлечение описаний товаров и услуг
- Мониторинг новостей, погоды и объявлений
- Сбор данных для аналитики
- Сбор данных для машинного обучения

Почему Python отлично подходит для веб-скрейпинга?

- Простой синтаксис
- Мощные библиотеки
- Быстрое выполнение
- Обширная документация



Типы HTTP-запросов

GET - запрашивает представление ресурса. Запросы с использованием этого метода могут только извлекать данные.

HEAD - запрашивает ресурс так же, как и метод GET, но без тела ответа.

POST - используется для отправки сущностей к определённому ресурсу. Часто вызывает изменение состояния или какие-то побочные эффекты на сервере.

PUT - заменяет все текущие представления ресурса данными запроса.

DELETE - удаляет указанный ресурс.

PATCH - используется для частичного изменения ресурса.

Selenium

Selenium - это библиотека для автоматизации веб-браузера. Она позволяет программно управлять браузером: эмулировать действия реального пользователя в браузере

```
from selenium import webdriver

from selenium.webdriver.common.by import By

driver = webdriver.Chrome()

driver.get("https://www.example.com")

heading = driver.find_element(By.TAG_NAME, "h1")

print(heading.text)

driver.quit()
```